

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA BIOMEDICÍNSKÉHO INŽENÝRSTVÍ
Katedra biomedicínské techniky



Automatická klasifikace segmentů EEG signálu za pomoci
algoritmu DENCLUE

Automatic classification of EEG segments using DENCLUE
algorithm

Diplomová práce

Vedoucí práce: doc. Ing. Vladimír Krajča, CSc.

Student: Bc. Jan Štrobl

květen 2016

Katedra biomedicínské techniky

Akademický rok: 2015/2016

Zadání diplomové práce

Student: **Bc. Jan Štrobl**
Studijní obor: Biomedicínský inženýr
Téma: **Automatická klasifikace segmentů EEG signálu za pomoci algoritmu DENCLUE**
Téma anglicky: Automatic classification of EEG segments using DENCLUE algorithm

Zásady pro vypracování:


Jedním z problémů automatického zpracování EEG záznamu je klasifikace již segmentovaného záznamu. Použijte metodu DENCLUE pro klasifikaci dat na základě příznaků. V programovém prostředí MATLAB implementujte uživatelsky přívětivé prostředí pro načtení analyzovaných dat. Implementujte algoritmus DENCLUE pro automatickou klasifikaci EEG signálu na základě příznaků. Testujte algoritmus DENCLUE na simulovaných datech. Proveďte kvalitativní i kvantitativní zhodnocení použitého algoritmu. Aplikujte algoritmus DENCLUE na reálné EEG záznamy pacientů. Proveďte kvalitativní i kvantitativní zhodnocení použitého algoritmu. Porovnejte a statisticky vyhodnoťte rozdíl v účinnosti algoritmu DENCLUE a k-means u simulovaného i reálného EEG signálu. Analyzujte výpočetní náročnost obou algoritmu.

Seznam odborné literatury:

- [1] Krajča V., Mohylová J., Číslíkové zpracování neurofyzilogických signálů, ed. Fakulta biomedicínského inženýrství, ČVUT Praha, 2011, ISBN 978-80-01-04721-7
- [2] Alexander Hinneburg, Daniel A. Keim, A General Approach to Clustering in Large Databases with Noise, Knowledge and Information Systems, ročník 5, číslo 4, 2003, 387-415 s.
- [3] Proakis J.G., Manolakis D.G., Digital Signal Processing, ed. 4th, Macmillan, NY, 2006, ISBN 978-0131873742

Vedoucí: **doc. Ing. Vladimír Krajča, CSc.**
Konzultant: **Ing. Václava Sedlmajerová**

Zadání platné do: 20.08.2017


.....
vedoucí katedry / pracoviště


.....
děkan

V Kladně dne 20.01.2015

Obsah

Seznam použitých symbolů a zkratek	3
Seznam tabulek	4
Seznam obrázků	7
Úvod	15
1 Teoretický základ problematiky	17
1.1 Elektroencefalografie	17
1.2 Automatická klasifikace EEG	21
1.2.1 Popis využívaných příznaků	22
1.2.2 Algoritmy a současný stav problematiky	26
1.3 Algoritmus DENCLUE	29
2 Metodologie	39
2.1 Programová specifikace	39
2.1.1 Wafe-Finder (WF)	40
2.1.2 Modul v MATLABu	43
2.2 Použitá data	44
2.2.1 Simulovaná data	44
2.2.2 Reálné EEG záznamy	47
2.3 Statistické vyhodnocení dat	47
2.3.1 Kvantitativní analýza	47
2.3.2 Kvalitativní analýza	47
2.4 Úprava algoritmu DENCLUE na EEG data	52
2.4.1 Nastavení parametrů	59
3 Výsledky	65
3.1 Simulovaná data	65
3.1.1 Kvantitativní analýza	65

3.1.2	Kvalitativní analýza	66
3.2	Reálné EEG záznamy	73
3.2.1	Kvantitativní analýza	74
3.2.2	Kvalitativní analýza celého EEG záznamu	76
3.2.3	Kvalitativní analýza pro více EEG záznamů	88
4	Diskuze	92
4.1	Simulovaná data	92
4.2	Reálné EEG záznamy	94
5	Závěr	100
	Reference	102
	Přílohy	106
A	Tabulky	106
B	CD	109

Seznam použitých symbolů a zkratek

1D	1-dimenzionální
2D	2-dimenzionální
ASH	Average Shifted Histograms (Průměr posunutých histogramů)
DBSCAN	Density Based Clustering
DENCLUE	Density Based Clustering
EEG	Elektroencefalografie
FN	False Negative (Falešně negativní)
FP	False Positive (Falešně pozitivní)
GUI	Grafické uživatelské rozhraní (Graphic User Interface)
KDE	Kernel Density Estimation (Jádrový odhad hustoty)
k-NN	k-Nearest Neighbors (k-nejbližších sousedů)
MATLAB	Matrix Laboration
PPV	Prediktivní hodnota pozitivního testu (Positive Prediktive Value)
ROC	Receiver Operating Characteristics
TN	True Negative (Správně negativní)
TP	True Positive (Správně pozitivní)
WF	Wave-Finder

Seznam tabulek

1	Seznam příznaků, které lze v programu Wave-Finder využít pro automatickou klasifikaci EEG záznamu.	23
2	Nastavení parametrů adaptivní segmentace pro tuto práci.	42
3	Mnou vytvořené funkce v programu MATLAB R2015a.	43
4	Časová náročnost u algoritmů DENCLUE a k-means pro počet segmentů simulovaných dat.	65
5	Specifická, senzitivita a PP pro algoritmy k-means a DENCLUE u simulovaných dat s výrazně oddělenými shluky.	68
6	Specifická, senzitivita a PP pro algoritmy k-means a DENCLEU u simulovaných dat s nevýrazně oddělenými shluky.	70
7	Specifická, senzitivita a PP pro algoritmy k-means a DENCLUE u simulovaných dat se shluky o různé vzájemné vzdálenosti a hustotě.	71
8	Specifická, senzitivita a PP pro algoritmy k-means a DENCLEU u simulovaných dat s prolnutými shluky.	73
9	Časová náročnost u testovaných algoritmů pro počet segmentů reálných EEG záznamů.	74
10	Počet segmentů pro každý hledaný shluk a celkový EEG záznam subjektu 1.	76
11	Specifická, senzitivita a PPV pro testované algoritmy a hledané shluky EEG záznamu subjektu 1.	77
12	Počet segmentů pro každý hledaný shluk a celkový EEG záznam subjektu 2.	79
13	Specifická, senzitivita a PPV pro testované algoritmy a hledané shluky EEG záznamu subjektu 2.	79
14	Počet segmentů pro každý hledaný shluk a celkový EEG záznam subjektu 3.	82
15	Specifická, senzitivita a PPV pro testované algoritmy a hledané shluky EEG záznamu subjektu 3.	82
16	Počet segmentů pro každý hledaný shluk a celkové EEG záznamy subjektů 1, 2 a 3.	85
17	Specifická, senzitivita a PPV pro testované algoritmy a hledané shluky EEG záznamů subjektů 1 až 3.	85

18	Počet segmentů u každého hledaného shluku z 50 náhodných segmentů EEG záznamu subjektu 4.	89
19	Příklad specificity, senzitivity a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 4.	89
20	Počet segmentů pro každý hledaný shluk 12 subjektů u nichž bylo náhodně vybráno 50 segmentů z EEG záznamu.	90
21	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamů 12 subjektů.	90
22	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 5.	106
23	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 6.	106
24	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 7.	106
25	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 8.	107
26	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 9.	107
27	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 10.	107
28	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 11.	108
29	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 12.	108
30	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 13.	108
31	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 14.	109
32	Specificita, senzitivita a PPV pro testované algoritmy a hledané shluky u náhodně vybraných segmentů EEG záznamu subjektu 15.	109

33	Seznam příloh na CD.	109
----	------------------------------	-----

Seznam obrázků

1	Příklady alfa, beta, theta, delta aktivity a epileptických grafoelementů v EEG záznamu.	19
2	Příklad svalového artefaktu v EEG záznamu.	20
3	Příklad technického artefaktu síťového brumu v EEG záznamu.	20
4	Rozvržení elektrod na hlavě pacienta během měření EEG systémem 10-20 a zkratkovité názvy měřicích elektrod.	21
5	Popis principu funkce algoritmu automatické klasifikace k-means.	28
6	Příklad čtvercové a Gaussovo jádrové funkce používaných v KDE.	30
7	Princip skládání celkové funkce hustoty z lokálních funkcí hustoty u algoritmu DENCLUE.	31
8	Příklad histogramu ve 2 dimenzích.	32
9	Příklad celkové funkce hustoty a rozřazení objektů do shluků u algoritmu DENCLUE.	33
10	Příklad celkové funkce hustoty pro koeficient $h = 0,2$	34
11	Příklad celkové funkce hustoty pro koeficient $h = 0,6$	35
12	Příklad celkové funkce hustoty pro koeficient $k = 50$	36
13	Příklad celkové funkce hustoty pro koeficient $k = 20$	36
14	Princip funkce parametru ξ u algoritmu DENCLUE.	37
15	Blokové schéma znázorňující v této práci užívanou automatickou klasifikaci EEG záznamu algoritmem DENCLUE.	40
16	Příklad prostředí v praxi používaného programu WF.	41
17	Příklad zobrazení jednoho shluku vytvořeného automatickou klasifikací EEG záznamu v programu WF.	42
18	Jednoduché GUI využívané uživateli pro spuštění algoritmu DENCLUE.	44
19	Simulovaná data s výrazně oddělenými shluky.	45
20	Simulovaná data s nevýrazně oddělenými shluky.	45
21	Simulovaná data se shluky o různé vzájemné vzdálenosti a hustotě.	46
22	Simulovaná data s prolnutými shluky.	46
23	Příklad segmentů typicky zastupujících shluk fyziologické aktivity mozku.	49

24	Příklad segmentů typicky zastupujících shluk epileptických grafoelementů.	50
25	Příklad segmentů typicky zastupujících shluk pomalých očních artefaktů.	50
26	Příklad segmentů typicky zastupujících shluk EMG artefaktů a zašuměného signálu.	50
27	Příklad segmentů typicky zastupujících shluk rovných čar.	50
28	Příklad segmentů typicky zastupujících shluk impulzních artefaktů.	50
29	Příklady špatně nasegmentovaných segmentů.	51
30	Příklad celkové funkce hustoty pro příznaky „maximální pozitivní amplituda v daném segmentu“ a „maximální negativní amplituda v daném segmentu“.	53
31	Příklad celkové funkce hustoty pro příznaky „maximální hodnota 1. derivace v segmentu“ a „maximální hodnota 2. derivace v segmentu“.	54
32	Příklad celkové funkce hustoty pro příznaky „Hjorthův parametr aktivity“ a „nelineární energie segmentu“.	54
33	Příklad celkové funkce hustoty pro příznaky „střední hodnota první derivace v segmentu“ a „maximální hodnota první derivace v segmentu“ a na jejím základě barevně oddělené dva shluky.	56
34	Příklad celkových funkcí hustoty pro příznaky „maximální pozitivní amplituda v daném segmentu“ a „maximální negativní amplituda v daném segmentu“ před a po odebrání nejvýraznějšího shluku.	57
35	Příklad celkových funkcí hustoty pro příznaky „maximální hodnota 1. derivace v segmentu“ a „maximální hodnota 2. derivace v segmentu“ před a po odebrání nejvýraznějšího shluku.	58
36	Graf závislosti počtu shluků na koeficientu ξ .	60
37	Příklad rozdělení objektů do shluků u algoritmu DENCLUE a koeficientu $\xi = 10$.	60
38	Příklad rozdělení objektů do shluků u algoritmu DENCLUE a koeficientu $\xi = 100$.	61
39	Graf závislosti počtu shluků na koeficientu k .	62

40	Příklad rozdělení objektů do shluků u algoritmu DENCLUE a koeficientu $k = 10$	63
41	Příklad rozdělení objektů do shluků u algoritmu DENCLUE a koeficientu $k = 40$	63
42	Graf závislosti časové náročnosti algoritmů k-means a DENCLUE na počtu segmentů simulovaných dat.	66
43	Rozřazení simulovaných dat s výrazně oddělenými shluky u algoritmů DENCLUE a k-means	67
44	Rozřazení simulovaných dat s nevýrazně oddělenými shluky u algoritmů DENCLUE a k-means	69
45	Rozřazení simulovaných dat shluků o různé vzájemné vzdálenosti a hustotě u algoritmů DENCLUE a k-means	71
46	Rozřazení simulovaných dat s prolnutými shluky u algoritmů DENCLUE a k-means	72
47	Graf závislosti časové náročnosti algoritmu DENCLUE neupraveného na EEG záznam na počtu segmentů reálných EEG záznamů.	75
48	Graf závislosti časové náročnosti algoritmů k-means a DENCLUE upraveného na EEG záznam na počtu segmentů reálných EEG záznamů.	75
49	Příklad shluku epileptických grafoelementů vytvořeného algoritmem k-means z EEG záznamu subjektu 1.	78
50	Příklad shluku EMG artefaktů a zašuměného signálu vytvořeného algoritmem DENCLUE upraveným na EEG záznam z EEG záznamu subjektu 1.	78
51	Příklad shluku rovných čar vytvořeného algoritmem DENCLUE upraveným na EEG záznam. Všechny segmenty patří do shluku. Příklad je vzat z EEG záznamu subjektu 2. Zobrazeno v programu WF [5].	80
52	Příklad shluku impulzních artefaktů vytvořeného algoritmem k-means z EEG záznamu subjektu 2.	81
53	Příklad shluku EMG artefaktů a zašuměného signálu vytvořeného algoritmem k-means z EEG záznamu subjektu 3.	83

54	Příklad shluku EMG artefaktů a zašuměného signálu vytvořeného algoritmem DENCLUE upraveným na EEG záznam z EEG záznamu subjektu 3.	84
55	Příklad shluku pomalých očních artefaktů vytvořeného algoritmem k-means z EEG záznamů subjektů 1 až 3.	86
56	Příklad shluku epileptických grafoelementů vytvořených algoritmem k-means z EEG záznamů subjektů 1 až 3.	87
57	Příklad shluku epileptických grafoelementů vytvořených algoritmem k-means z EEG záznamů subjektů 1 až 3.	87
58	Příklad shluku fyziologické aktivity mozku vytvořeného algoritmem DENCLUE upraveným na EEG záznam z EEG záznamů subjektů 1 až 3.	88

Abstrakt

EEG (elektroencefalografie) měří elektrický potenciál z povrchu hlavy. Z důvodu stochastické povahy je EEG záznam obtížně vyhodnocovatelný. S detekcí EEG záznamu mohou expertovi pomoci algoritmy automatické klasifikace. V rámci rešerše jsem nenalezl studie, které by ukazovaly stoprocentní úspěšnost rozřazení všech klinicky významných částí EEG záznamu bez použití trénovacích vzorů. V této práci jsem se zaměřil na ověření použitelnosti hustotního algoritmu DENCLUE.

Pro segmentaci signálu, extrakci příznaků a vizualizaci EEG jsem používal program Wave-Finder. Klasifikaci algoritmu DENCLUE jsem porovnával s klasifikací algoritmem k-means (funkce programu Wave-Finder). V programovém prostředí MATLAB R2015a jsem vytvořil modul pro klasifikaci algoritmem DENCLUE.

Kvalitativně (ROC analýza) a kvantitativně jsem vyhodnotil použité algoritmy na simulovaných datech a na reálných EEG záznamech. Kvalitativní vyhodnocení EEG záznamů prováděli 2 experti. EEG záznamy byly naměřeny v Nemocnici Na Bulovce na 15 subjektech. Jednalo se o subjekty různého pohlaví ve věku 26 až 60 let.

Výsledky ukazují, že algoritmus DENCLUE dokázal dobře rozeznat část EEG záznamu (například fyziologickou aktivitu mozku). Algoritmus ovšem nebyl schopný dobře rozeznat tu část EEG záznamu, která je pro lékaře podstatnější (například epileptické grafoelementy). Z tohoto důvodu není prozatím algoritmus DENCLUE vhodný na automatickou klasifikaci EEG záznamu. Jeho vhodnost by se mohla vylepšit extrahováním jiných příznaků.

Klíčová slova

DENCLUE, EEG, automatická klasifikace, k-means, ROC analýza

Abstract

EEG (electroencephalography) measures the electrical potential on the scalp. EEG recordings have a stochastic nature, their evaluating is difficult. Automatic classification algorithms can prove being useful with EEG detection. To my knowledge, there is no study that shows one hundredpercent successful rate for the profiling of all clinically significant parts of EEG patterns and wasn't using of training patterns. In this work I focused on verifying the possible applicability of the density algorithm DENCLUE.

I used the Wave-Finder program for segmentation of the signal, feature extraction and visualization of EEG. I compared classification capability of DENCLUE algorithm with the classification capability of k-means algorithm (Wave-Finder contains k-means). I created a module for classification algorithm DENCLUE in MATLAB R2015a .

I have evaluated used algorithms on simulated data and real EEG recordings qualitatively (ROC analysis) and quantitatively. Qualitative evaluation of EEG were performed by two experts. EEGs were recorded in hospital Na Bulovce on 15 subjects. These subjects were of different sex and ages range was between 26 - 60 years old.

The results showed that the algorithm DENCLUE is able to clearly discern parts of EEG. However, the algorithm is not able to separate parts of the EEG, which is fundamental for doctors (such as epileptic graphoelements). For this reason DENCLUE algorithm is not yet suitable for the automatic classification of the EEG. Its suitability could be improved the extraction of other features.

Keywords

DENCLUE, EEG, automatic classification, k-means, ROC analysis

Poděkování

Rád bych poděkoval svému vedoucímu doc. Ing. Vladimíru Krajčovi, CSc. za jeho příkladné vedení po celou dobu vzniku této práce a za jeho kvalitativní hodnocení EEG záznamů. Za kvalitativní hodnocení EEG záznamů bych rád poděkoval i primáři MUDr. Ing. Svojmilu Petránkovi, MBA. Dále bych rád poděkoval svojí konzultantce Ing. Václavě Sedlmajerové za značnou ochotu v období konzultací. Za konzultaci statistického vyhodnocení výsledků práce bych rád poděkoval doc. Ing. Janě Vránové, CSc. V neposlední řadě bych rád poděkoval všem členům EEG týmu Fakulty biomedicínského inženýrství ČVUT v Praze.

Prohlášení

Prohlašuji, že jsem diplomovou práci s názvem

.....

vypracoval samostatně a použil k tomu úplný výčet citací použitých pramenů, které uvádím v seznamu přiloženém k diplomové práci.

Nemám závažný důvod proti užití tohoto školního díla ve smyslu § 60 Zákona č.121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V dne

.....

Podpis

Úvod

EEG snímá elektrický potenciál mozku. Jeho využití je velmi významné pro určování epilepsie. Mezi další využití patří například detekce fází spánku či hledání ložiskových abnormalit [1, 2]. EEG záznam má stochastickou a nestacionární povahu. V klinické praxi se často měří dlouhodobý EEG záznam. Tyto faktory mají za následek složité vyhodnocování záznamu i pro zkušeného experta [3]. S vyhodnocováním EEG záznamu by měly expertovi pomoci algoritmy automatické klasifikace.

Algoritmy automatické klasifikace jsou stále populárnější například pro detekci epilepsie [4]. Tyto algoritmy rozdělují data do skupin podle jejich charakteristiky. Data mající stejnou charakteristiku jsou zařazována do stejné skupiny. Pro správnou funkci algoritmů automatické klasifikace je potřeba nejprve provést segmentaci a extrakci příznaků. Segmentace rozdělí vyhodnocovaný záznam na jednotlivé segmenty. V této práci budu využívat adaptivní segmentaci vytvořenou programem, jenž je pro tento účel využíván v klinické praxi. Konkrétně se bude jednat o program Wave-Finder (WF), který vytvořil doc. Ing. Vladimír Krajča, CSc. [5]. Každý ze vzniklých segmentů je charakterizován za pomoci extrakce příznaků. Algoritmy automatické klasifikace jsou silně závislé na použitých příznacích [6]. Existuje řada studií využívajících příznaky vybírané pro EEG záznam (viz například [6], [7] nebo [8]). V této práci budu využívat pouze příznaky získávané z programu WF.

Existuje několik různých algoritmů automatické klasifikace. Jejich základní dělení je na algoritmy s učitelem a bez učitele. Pro použití algoritmů s učitelem je potřeba mít správné trénovací vzory [9]. Cílem práce je implementovat na EEG záznam algoritmus automatické klasifikace nepotřebující trénovací vzory čili algoritmus bez učitele. Jedním z nejčastěji využívaných algoritmů bez učitele je algoritmus k-means [10]. Mezi nevýhody algoritmu k-means patří nemožnost automatického určení počtu shluků, velká citlivost na iniciační centra shluků a neschopnost klasifikovat prolnuté shluky [11]. Tyto nedostatky neplatí pro algoritmus automatické klasifikace DENCLUE (Density Based Clustering) [12].

Algoritmus DENCLUE je hustotně založený algoritmus automatické klasifikace. Na rozdíl od jiných algoritmů na bázi hustoty typu DBSCAN je algoritmus DENCLUE založen na pravděpodobnostním rámci [13]. Základ algoritmu DENCLUE tvoří jádrový odhad hustoty (KDE - Kernel Density Estimation). Parametry algoritmu DENCLUE jsou koeficient šumu ξ a koeficient hladkosti h . V době vzniku této práce jsem nenalezl žádné studie zabývající se adaptací algoritmu DENCLUE na EEG záznam.

Adaptací algoritmu DENCLUE na EEG záznam se budu zabývat v této práci. Pro ověření jeho funkčnosti využiji porovnání s algoritmem automatické klasifikace používaným v klinické praxi. Tímto algoritmem bude algoritmus k-means z programu WF. Pro oba algoritmy nejprve provedu kvalitativní a kvantitativní statistické hodnocení na simulovaných datech. Následně provedu adaptaci algoritmu DENCLUE na EEG data. Poté pro něj vytvořím grafické uživatelské rozhraní. Adaptovaný a neadaptovaný algoritmus DENCLUE spolu s algoritmem k-means budu kvalitativně a kvantitativně porovnávat. Toto porovnání provedu pro alespoň 3 celé EEG záznamy. Dále zjistím účinnost upraveného algoritmu DENCLUE a algoritmu k-means u více naměřených subjektů.

Cílem této práce je aplikovat algoritmus DENCLUE na EEG záznam a pomocí statistického vyhodnocení porovnat jeho účinnost na automatickou klasifikaci EEG záznamů a simulovaných dat s algoritmem k-means používaným v klinické praxi.

1 Teoretický základ problematiky

Celá následující práce se zabývá zpracováním biologického signálu, konkrétně elektrického signálu získaného pomocí EEG. EEG signál patří mezi signály stochastické a nestacionární povahy [3]. Jedná se o náhodný signál, jenž nelze popsat explicitními matematickými vztahy. Charakteristika EEG signálu je popsána pomocí statistických parametrů. Tyto parametry nejsou funkcí času [14]. Dlouhodobý signál charakterizovaný statistickými parametry se stává zdlouhavým a náročným na vyhodnocení i pro zkušeného experta [3]. Usnadnění práce experta by měla zajistit automatická klasifikace EEG záznamu, která je čím dál více populární například pro predikci epilepsie [4].

V této kapitole je posána samotná elektroencefalografie, následně automatická klasifikace EEG záznamu a algoritmus automatické klasifikace DENCLUE, jehož funkčnost na automatickou klasifikaci EEG záznamu v této práci ověřuji.

1.1 Elektroencefalografie

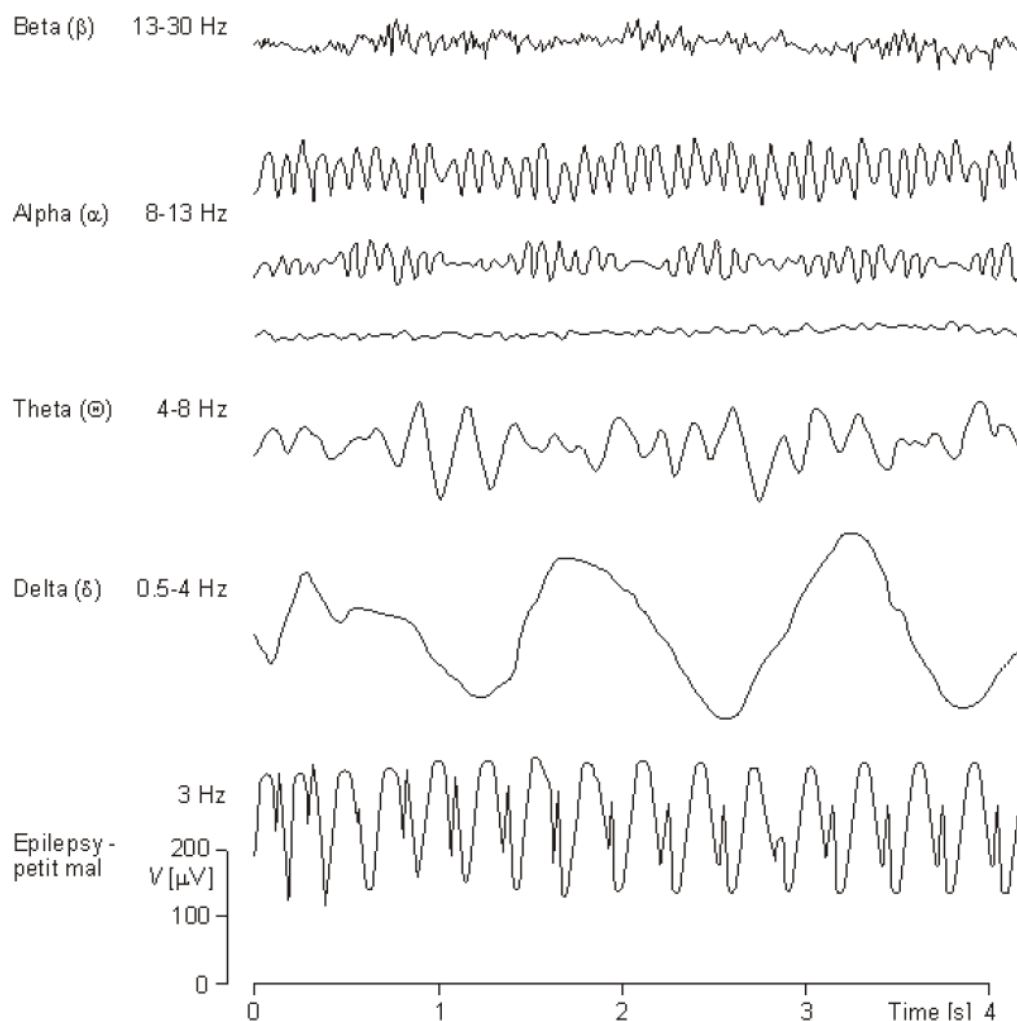
Elektroencefalografie vytváří záznam elektrických potenciálů mozku, takzvaný elektroencefalogram. Snímání záznamu probíhá za pomoci elektrod umístěných u běžného EEG na povrchu hlavy měřeného subjektu [1, 2]. Přístroj, jenž tuto aktivitu snímá se nazývá elektroencefalograf.

Základní stavební jednotkou mozku jsou buňky zvané neurony, které přenáší elektrický potenciál. Typický neuron se skládá z těla (soma), odstředivých drah (dendrity) a odstředivé dráhy (axonu). Klasické EEG měří sumační elektrický signál z neuronů umístěných v povrchových vrstvách mozkové kůry. Povrchová vrstva mozkové kůry je část mozku nejbližší měřicím elektrodám. Mezi somatem a dendrity neuronů této vrstvy se vytváří stále se měnící dipól [1]. Lidskou tkáň v okolí těchto neuronů lze považovat za objemový vodič, což způsobuje, že se v ní stále se měnící dipól projevuje jako vlnovitě se měnící potenciál [1]. Elektrody zaznamenávají rytmický signál pouze pokud je aktivita neuronů synchronizována. V opačném případě se různé fáze elektrických potenciálů neuronů vyruší. Na synchronizaci měřeného signálu má vliv thalamus, který plní funkci generátoru rytmů [10]. Normální EEG aktivita je většinou rytmická a má sinusoidní cha-

rakter [10]. Amplituda signálu je zmenšená, neboť se zaznamenává přes lebeční kosti. [10, 1]

V EEG záznamu lze rozeznat alfa, beta, delta a theta aktivitu (viz obrázek 1). Aktivity se od sebe odlišují svojí amplitudou a frekvencí. Alfa aktivita je fyziologicky dominantní až u zralého mozku [10]. Alfa aktivita se fyziologicky vyskytuje převážně v parieto-okcipitální oblasti při bdělém a relaxovaném stavu měřeného subjektu. Relaxovaný subjekt má v tomto stavu zavřené oči, ovšem nespí. Pokud měřený subjekt otevře oči, dochází k takzvané desynchronizaci měřeného signálu a alfa aktivita se ztrácí. Alfa aktivita je charakteristická frekvencí 8 - 13 Hz a amplitudou 50 - 100 μV . Ve frontální oblasti je při stejných podmínkách dominantní beta aktivita. Beta aktivita se vyznačuje frekvencí od 13 do 30 Hz a amplitudou většinou menší než je 30 μV . Mezi další významné fyziologické rytmy patří delta aktivita a theta aktivita. Delta aktivita má frekvenci pod 4 Hz a amplitudu 10 - 300 μV . Fyziologicky se delta aktivita vyskytuje u dětí do 3 let a při hlubokém spánku. Theta aktivita má typickou frekvenci 4 až 8 Hz a amplitudu 30 - 50 μV . Vyskytuje se u dětí v bdělém stavu. Významná je i gama aktivita s nízkou amplitudou. Objevuje se při soustředěné pozornosti u zralého mozku a má frekvenci 30 až 80 Hz. Obecně se dá říci, že čím větší je frekvence rytmu, tím menší je jeho amplituda. Na obrázku 1 jsou ukázány alfa, beta, theta, delta aktivity a jejich charakteristiky. [10, 1]

Další charakteristické aktivity v EEG záznamu se nazývají grafoelementy. Významné jsou epileptické grafoelementy. Zobrazují se jako ostré vlny, hroty nebo komplexy hrotů následovaných pomalou vlnou [2] (viz obrázek 1). Tyto grafoelementy se využívají při detekci epilepsie. Epilepsie je onemocnění vznikající z různých příčin. Vyznačuje se opakovanými nervovými záchvaty, mezi nimiž je pacient obvykle bez výrazných obtíží. Epilepsie může být typu grand mal doprovázená ztrátou vědomí, tonickými a klonickými křečemi. Dalším typem epilepsie je patit mal. Při epilepsii typu patit mal pacient dočasně ztrácí kontakt s okolím. Jednotlivé druhy epileptických záchvatů mají svůj charakteristický EEG signál. [1, 10]



Obrázek 1: Příklady alfa, beta, delta a theta aktivity viditelné na EEG záznamu a charakteristické frekvence zobrazených grafoelementů. Příklad epileptického grafoelementu při záchvatu patit mal, viditelný na EEG záznamu a jeho charakteristická frekvence [14].

V EEG signálu se vyskytují i takzvané artefakty. Artefakt je naměřený elektrický potenciál, který nepochází ze sumačního potenciálu korových neuronů. Artefakty dělíme do dvou skupin. První skupinou jsou biologické artefakty, druhou skupinou jsou technické artefakty. Biologické artefakty mají biologickou povahu vzniku, ovšem jejich geneze je z jiného orgánu nežli mozku. Mezi biologické artefakty patří například svalová aktivita měřeného subjektu (viz obrázek 2) či artefakty vzniklé otevřením a zavřením očí. Technické artefakty nemají svůj původ v biologické aktivitě. Mezi technické artefakty patří

například síťový brum o frekvenci 50 Hz (viz obrázek 3) nebo artefakty vzniklé chybnou měřicí elektrodou. [10, 14]

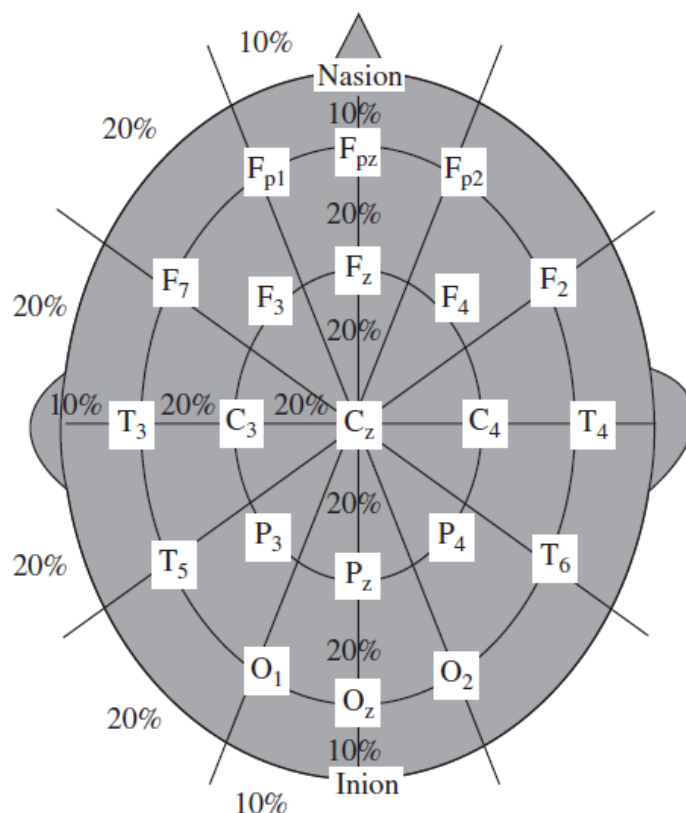


Obrázek 2: Příklad biologického artefaktu EEG záznamu, konkrétně svalového artefaktu. Svalový artefakt je červeně zakroužkován. Převzato z [10], upraveno.



Obrázek 3: Příklad technického artefaktu EEG záznamu, konkrétně síťového brumu [10].

Klasické EEG je snímáno elektrodami z povrchu hlavy. Běžně se používá 19 měřicích elektrod, umístěných podle systému 10-20. Tento systém standardizuje umístění elektrod. Elektrody jsou od sebe vzdálené 10 % a 20 % v každé rovině (viz obrázek 4, na němž jsou jednotlivé elektrody pojmenovány. Oproti běžnému použití jsou zde navíc elektrody F_{pz} a O_z). Méně časté je použití 64 měřicích elektrod, zapojených podle systému 10-10 [10]. Elektrický signál může být měřen vzhledem k jedné referenční elektrodě (unipolární zapojení elektrod), popřípadě jako rozdíl elektrického potenciálu mezi jednotlivými měřicími elektrodami (bipolární zapojení elektrod). U unipolárního zapojení se jako referenční elektroda bere buď elektroda z ušního lalůčku, nebo tzv. Goldmannova elektroda. Goldmanova elektroda je vytvořena spojením všech měřicích elektrod přes odpor k zemi. Bipolární zapojení se provádí buď transverzálně (příčně k ose hlavy), nebo longitudiálně (podélně s osou hlavy). Z unipolárního zapojení se obvykle snímá devatenácti kanálové EEG, kdežto z bipolárního zapojení se snímá zpravidla šestnácti kanálové EEG. [10]



Obrázek 4: Rozvržení elektrod na hlavě pacienta během měření EEG systémem 10-20 a zkratkovité názvy měřicích elektrod [15].

EEG má velký význam, při detekci epilepsie. Epilepsií trpí v Evropě a USA přibližně 1 % populace [10]. Mezi další použití EEG patří například využití v oblasti poruch spánku [1] či hledání ložiskových i difuzních abnormalit v mozku. Tyto abnormality se projevují smazáním nebo rozpadem rytmu a zpomalením do theta nebo delta pásma [2].

1.2 Automatická klasifikace EEG

Hlavním cílem automatické klasifikace je rozdělení jednotlivých částí signálu do homogenních skupin. Výsledkem automatické klasifikace jsou skupiny, v nichž jsou pouze části analyzovaného signálu s podobnými charakteristikami. [16]

Rozdělení signálu na části provádí segmentace signálu. Existují dva typy segmentace, fixní a adaptivní. Fixní segmentace rozděluje signál do stejně dlouhých částí (segmentů). Fixní segmentace může způsobit vznik segmentu obsahujícího například část alfa akti-

vity a část epileptického grafoelementu. Tyto dvě části by měly být od sebe oddělené. Adaptivní segmentace vytváří různě dlouhé segmenty podle charakteru signálu. Adaptivní segmentaci jsem využíval i v této práci jako předfázi aplikace jednotlivých algoritmů automatické klasifikace. [16]

Po segmentaci signálu je možné jednotlivé segmenty klasifikovat. Rozdělení segmentů se provádí do skupin (shluků). Segmenty mající stejný charakter patří do stejného shluku. Toto rozdělení provádí algoritmy automatické klasifikace na základě vypočtených charakteristik segmentů. Tyto charakteristiky se nazývají příznaky. Příkladem používaného příznaku může být maximální amplituda signálu v segmentu či průměrná první derivace signálu v daném segmentu. Extrakce příznaků je výpočet příznaků pro každý segment zvlášť. Algoritmy automatické klasifikace pracují v příznakovém prostoru, jehož dimenze se rovná počtu počítaných příznaků. V příznakovém prostoru se používá termín objekty namísto termínu segmenty.

Existuje velké množství příznaků využívaných pro automatickou klasifikaci EEG signálu. Vliv výběru správné skupiny příznaků na funkčnost metod automatické klasifikace je výrazný [6]. Velké množství studií se zabývá výběrem správných příznaků pro automatickou klasifikaci EEG signálu. Například Georgiy Minasyan ve své studii (viz [6]) klasifikuje epileptické záchvaty za pomoci výběru z 58 příznaků. Naopak například Gotman (viz [8]) používá pro detekci nástupu epileptického záchvatu pouze 5 příznaků. Počet příznaků se stále zvětšuje o nově používané příznaky (viz například [7]). [16]

Pro zjištění využitelnosti mnou testovaného algoritmu automatické klasifikace bylo zapotřebí porovnání s algoritmem využívaným v klinické praxi. Porovnávacím algoritmem byl k-means využívaný v programu WF [5]. Ve své práci využívám pouze příznaky, které jsou implementovány v programu WF (viz kapitola 2.1). Výběr této sady příznaků odstraní případnou chybu, která by mohla vzniknout z používání různých příznaků pro dva navzájem porovnávané algoritmy.

1.2.1 Popis využívaných příznaků

V programu WF umožňuje extrakci 24 příznaků. Algoritmus k-means implementovaný v programu WF je využíváno prvních 23 příznaků z dané sady. Tyto příznaky byly ověřeny

v lékařské praxi. Všechny příznaky programu WF jsou normované. Normalizace je provedena podle maximální hodnoty daného příznaku. Příznaky tak nabývají hodnot od 0 do 1. Seznam všech 24 příznaků lze nalézt v tabulce 1. [16, 5]

Tabulka 1: Seznam příznaků, které lze v programu Wave-Finder využít pro automatickou klasifikaci EEG záznamu. Z těchto příznaků jsem vybíral příznaky pro tuto práci [5].

pořadí příznaku	název příznaku
1	variabilita signálu v daném segmentu
2	maximální pozitivní amplituda v daném segmentu
3	maximální negativní amplituda v daném segmentu
4	hodnota signálu v 1. části delta frekvenčního pásma (0,5 - 1,5 Hz)
5	hodnota signálu v 2. části delta frekvenčního pásma (2 - 3,5 Hz)
6	hodnota signálu v 1. části theta frekvenčního pásma (4 - 5,5 Hz)
7	hodnota signálu v 2. části theta frekvenčního pásma (6 - 7,5 Hz)
8	hodnota signálu v 1. části alfa frekvenčního pásma (8 - 10 Hz)
9	hodnota signálu v 2. části alfa frekvenčního pásma (10,5 - 12,5 Hz)
10	hodnota signálu v sigma frekvenčním pásmu (18 - 29 Hz)
11	hodnota signálu v beta frekvenčním pásmu (13,5 - 17,5 Hz)
12	maximální hodnota první derivace v segmentu
13	maximální hodnota druhé derivace v segmentu
14	hodnota střední frekvence v segmentu
15	střední hodnota první derivace v segmentu
16	střední hodnota druhé derivace v segmentu
17	Hjorthův parametr mobility
18	Hjorthův parametr komplexity
19	Hjorthův parametr aktivity
20	délka křivky v segmentu
21	nelineární energie segmentu
22	počet průchodů signálu nulou
23	frekvence dominantní složky výkonu spektra
24	inflexní bod

„Maximální amplituda segmentu v negativním směru“ a „maximální amplituda segmentu v pozitivním směru“ značí maximální hodnotu napětí daného segmentu po odečtení stejnosměrné složky napětí. Výpočet stejnosměrné složky lze vidět v rovnici 1. Pozitivním směrem se u signálu EEG myslí záporná amplituda a opačně. [16]

$$A_{DC} = \frac{\sum_{i=1}^L x_i}{L}, \quad (1)$$

kde A_{DC} je stejnosměrná složka amplitudy signálu v daném segmentu, L je délka segmentu a x_i je amplituda i -tého vzorku v segmentu [16].

Mezi příznaky využívané programem WF patří dále „variabilita signálu v daném segmentu“. Dalšími příznaky jsou „maximální hodnota první derivace“ a „střední hodnota první derivace“. První derivace určuje sklon křivky grafoelementu a její výpočet lze aproximovat podle vzorce 2. [16]

$$\frac{dx}{dt} = x_{i+1} - x_i, \quad (2)$$

kde $\frac{dx}{dt}$ je derivace amplitudy vzorku podle času a x_i je amplituda i -tého vzorku v segmentu [16].

Příznaky „maximální hodnota druhé derivace“ a „střední hodnota druhé derivace“ odpovídají špičatosti křivky a lze je aproximovat podle rovnice 3. [16]

$$\frac{d^2x}{dt^2} = x_{i+4} - 2x_{i+2} + x_i, \quad (3)$$

kde $\frac{d^2x}{dt^2}$ je druhá derivace amplitudy vzorku podle času a x_i je amplituda i -tého vzorku v segmentu [16].

Významnými příznaky v automatické klasifikaci signálu EEG jsou tři Hjorthovy parametry. První Hjorthův parametr se nazývá aktivita. Aktivita se počítá podle vzorce 4 a popisuje výkon signálu v daném segmentu. Druhým Hjorthovým parametrem je mobilita. Mobilita je popsána rovnicí 5 a ukazuje střední hodnotu frekvence v segmentu. Třetí Hjorthův parametr se nazývá komplexita (viz rovnice 6) a popisuje podobnost signálu s čistě sinusovou vlnou. [17]

$$Activity = \sigma(x)^2, \quad (4)$$

kde *Activity* je Hjorthův parametr aktivity, $\sigma()$ označuje směrodatnou odchylku signálu v segmentu a x je amplituda vzorku v segmentu [16].

$$Mobility = \sqrt{\frac{Activity\left(\frac{dx(t)}{dt}\right)}{Activity(x(t))}}, \quad (5)$$

kde *Mobility* je Hjorthův parametr mobility, *Activity*() je Hjorthův parametr aktivity, $\frac{dx(t)}{dt}$ je derivace amplitudy vzorku segmentu podle času a $x(t)$ je amplituda vzorku v segmentu [17].

$$Complexity = \frac{Mobility\left(\frac{dx(t)}{dt}\right)}{Mobility(x(t))}, \quad (6)$$

kde *Complexity* je Hjorthův parametr komplexity, *Mobility*() je Hjorthův parametr mobility, $\frac{dx(t)}{dt}$ je derivace amplitudy vzorku segmentu podle času a $x(t)$ je amplituda vzorku v segmentu [17].

„Délka křivky segmentu“ se vypočítává v programu WF jako suma absolutních hodnot rozdílu amplitud jednotlivých vzorků v segmentu, jak lze vidět v rovnici 7. [16]

$$L = \sum_{i=1}^{N_s} abs[x(i) - x(i + 1)], \quad (7)$$

kde L je délka křivky signálu, N_s je počet vzorků v segmentu a $x(i)$ je amplituda i -tého vzorku v segmentu [16].

Příznak „počet průchodů signálu nulou“ určí kolikrát projde amplituda signálu bodem takzvané pozitivní nuly. Pozitivní nulou se myslí bod o velikosti elektrického napětí

$0,01 \mu\text{V}$. Příznak se vypočítává po odstranění stejnosměrné složky amplitudy (viz rovnice 1). Další příznak „nelineární energie segmentu“ se počítá podle vzorce 8. [16]

$$NLE(i) = x^2(i) - x(i-1)x(i+1), \quad (8)$$

kde $NLE(i)$ je nelineární energie i -tého vzorku segmentu a $x(i)$ je amplituda v i -tém vzorku segmentu [16].

Příznak „inflexní bod“ udává počet inflexních bodů v segmentu. „Hodnota střední frekvence v segmentu“ je příznak, jenž se získává za pomoci autokorelační funkce.

Program WF vypočítává i příznaky ze spektra signálu daného segmentu. Mezi takovéto příznaky patří „frekvence dominantní složky výkonu spektra“ a příznaky počítající hodnotu signálu v určité části frekvenčního spektra (příznaky jsou v tabulce 1 na straně 23 v pořadí 4., 5., 6., 7., 8., 9., 10. a 11.). Spektrum je získáno pomocí rychlé Fourierovy transformace. „Frekvence dominantní složky výkonu spektra“ je určena jako frekvence dominantního vrcholu spektra signálu. [16]

1.2.2 Algoritmy a současný stav problematiky

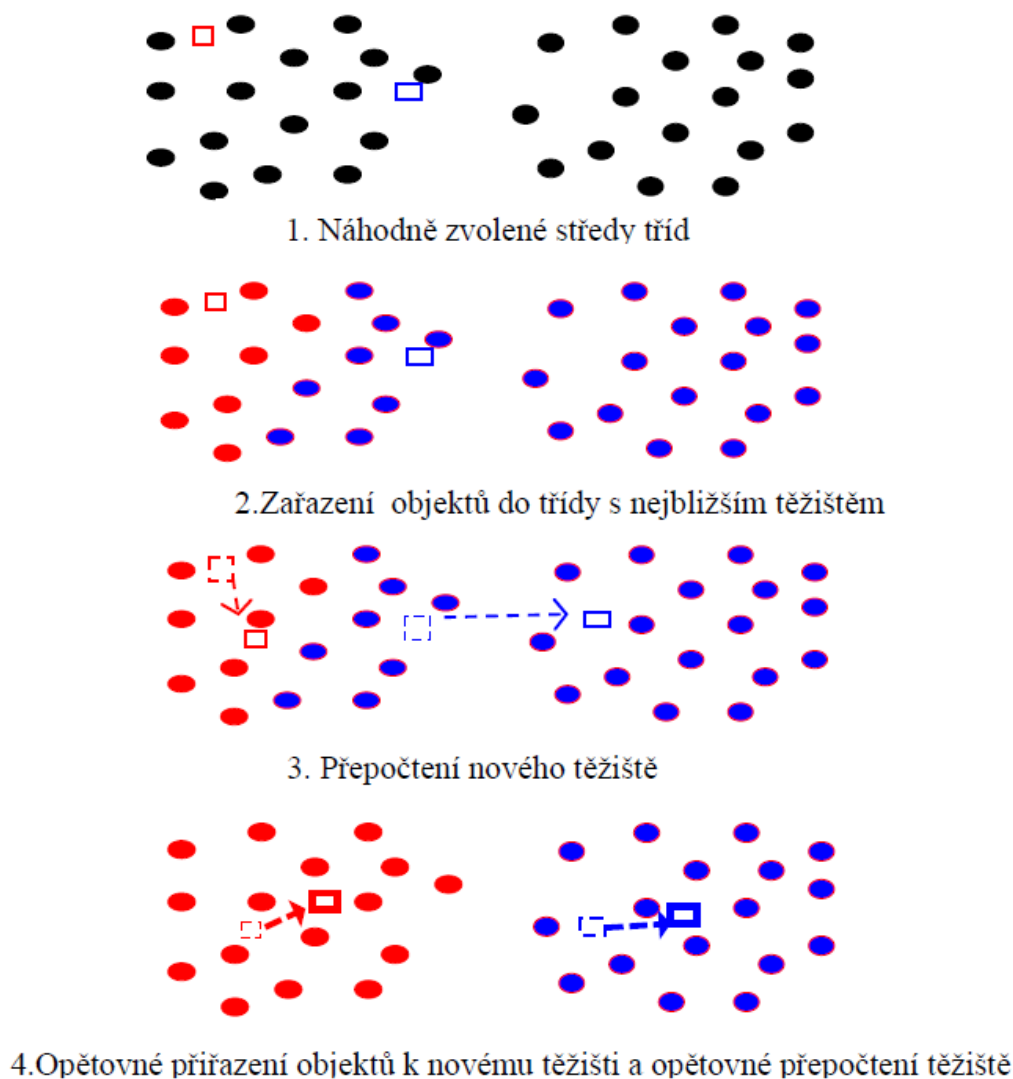
Existuje mnoho různých algoritmů automatické klasifikace. Tyto algoritmy se primárně dělí na učící se algoritmy a algoritmy bez učitele [16, 9]. Mezi učící se algoritmy patří například k-NN či neuronové sítě. Mezi klasifikátory bez učitele patří např. k-means, fuzzy c-means, hierarchické metody, DBSCAN, DENCLUE atd. [16, 10, 18]. Vstupním parametrem učících se klasifikátorů je trénovací množina dat. Na trénovací množině dat se učící klasifikátory naučí, jakým způsobem mají signál rozdělovat. Po fázi učení teprve dochází k automatickému rozdělování signálu [16, 9]. Pokud v klinické praxi není známá informace o měřených datech nebo neexistují trénovací vzory dat, nelze použít učící se klasifikátory. Problém nastává i pokud se v signálu vyskytují specifické segmenty, jejichž vzory nebyly vybrány do trénovací množiny dat [9]. Z těchto důvodů se ve své práci zaměřuji pouze na algoritmy automatické klasifikace bez učitele.

Mezi metodami automatické klasifikace bez učitele se nejčastěji pro automatickou klasifikaci EEG signálu používají metody shlukové analýzy. Především se používají paramet-

rická metoda k-means a neparametrické hierarchické metody [9]. Problém u hierarchických metod nastává při multikanálovém a dlouhodobém (řádově hodiny až dny) měření EEG. Důvodem problému je velká paměťová náročnost hierarchických metod [9]. Tato práce je zaměřena na zpracování multikanálového signálu EEG, jenž může být i dlouhodobý. Z těchto důvodů jsem jako porovnávací algoritmus využíval algoritmus k-means, zabudovaný v programu WF. Porovnáním s ním jsem zjišťoval účinnost mnou testovaného algoritmu.

Algoritmus k-means je všeobecně známá a v řadě různých aplikací často používaná metoda [10, 18, 19, 20, 21, 11]. Důvodem častého použití metody k-means je především její jednoduchost a efektivita [10, 9]. Při aplikaci na EEG signál bývá metoda k-means použita například pro detekci výskytu epileptických grafoelementů [20, 22], detekci fází spánku [23] a zjištění ospalosti řidiče [21]. V neposlední řadě je metoda k-means používána i v programu WF pro celkové rozdělení jednotlivých částí signálu podle jejich klinického významu [5].

Princip k-means spočívá v rozdělení jednotlivých objektů (zde segmentů signálu) do shluků podle vzdálenosti objektů od center shluků (viz obrázek 5). Nejprve se určí iniciální centra shluků (někdy též nazývaná těžiště) a následně se přidělí všechny objekty ke shluku s nejbližším centrem. Dále se přepočítají centra jednotlivých shluků a znovu se provede přiřazení objektů k shlukům s nejbližším centrem. K-means iteruje do té doby, dokud se mění zastoupení objektů v jednotlivých shlucích. Tento princip je sice jednoduchý a relativně účinný, ale má i řadu zásadních nevýhod. [16]



Obrázek 5: Popis principu funkce algoritmu automatické klasifikace k-means na dvou shlucích oddělených barvami (červená a modrá barva). Obdélníky představují centra shluků stejné barvy [16].

Mezi hlavní nevýhody algoritmu k-means patří nemožnost automatického určení počtu shluků [10, 11], což při špatném ručním nastavení vnáší chybu do klasifikace. Další nevýhodou je velká citlivost na určení iniciálního centra shluku [10, 11, 18, 19, 24, 25], což může vést ke konvergenci výsledku klasifikace do lokálního (nikoli globálního) minima vzdálenosti [10, 11, 25, 19]. Nevýhodou k-means je zároveň špatné klasifikování abnormálních dat [11]. Problém špatné klasifikace k-means pro některá data je dán již samotným principem metody. Ta je založená na principu porovnávání vzdálenosti shluků. Metoda tím pádem nemůže rozlišit dva shluky, které jsou prolnuté do sebe (viz obrázek 22

na straně 46). Tyto problémy se ovšem nevyskytují u řady metod automatické klasifikace bez učitele založené na bázi hustoty, mezi něž patří například DBSCAN či DENCLUE [18, 12, 13].

Tato práce je zaměřena na algoritmus DENCLUE, který původně vznikl na klasifikaci multimediálních databází. Tyto databáze jsou mimo jiné charakteristické velkým počtem příznaků (desítky až stovky příznaků) [12]. V tomto rozmezí příznaků se pohybuje i EEG záznam, který chceme klasifikovat (viz tabulka 1 na straně 23). Algoritmus DENCLUE se výrazně liší od ostatních algoritmů na bázi hustoty typu DBSCAN. Na rozdíl od nich používá DENCLUE pravděpodobnostní rámec [13].

Ačkoli existuje řada studií, které porovnávají jednotlivé algoritmy automatické klasifikace mezi sebou [18], nenalezl jsem žádné odborné studie, které by se zabývaly přímo porovnáním funkčnosti algoritmů DENCLUE a k-means. Zároveň jsem nenalezl studie zabývající se vyzkoušením algoritmu DENCLUE na klasifikaci EEG signálu.

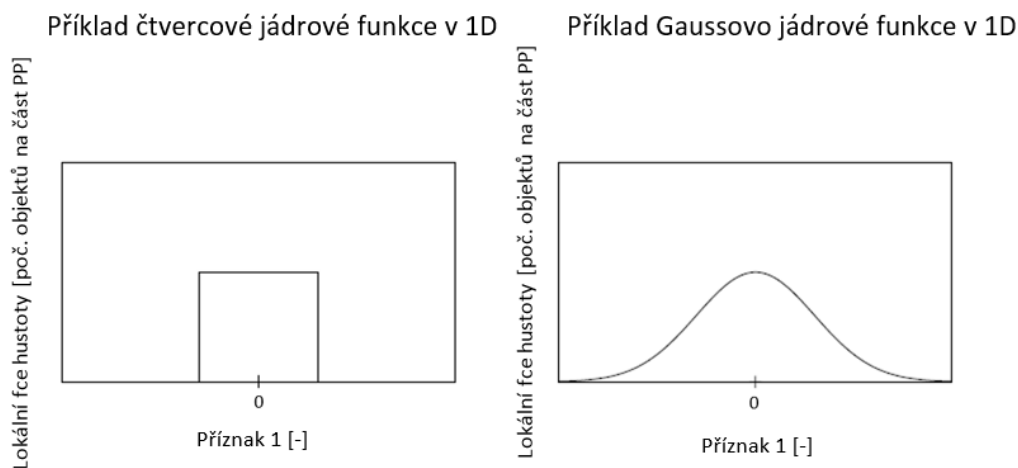
1.3 Algoritmus DENCLUE

Algoritmus DENCLUE je algoritmus automatické klasifikace založený na bázi hustoty. Na rozdíl od algoritmů automatické klasifikace na bázi hustoty typu DBSCAN je založen na matematickém základu. Tento matematický základ představuje statistická metoda KDE (Kernel Density Estimation - jádrový odhad hustoty). [12, 13]

KDE je statistická metoda, která popisuje vliv sousedních bodů (objektů) na počítaný objekt [12]. V KDE váhujeme pomocí jádrové funkce velikost vlivu jednotlivých objektů na výpočet hustoty v okolí počítaného objektu. Jádrová funkce určuje, jak moc tento vliv klesá se vzdáleností od počítaného objektu. Jádrové funkce mohou mít různý tvar. Mezi jádrové funkce patří například čtvercová nebo Gaussova jádrová funkce (viz obr. 6). V algoritmu DENCLUE se využívá Gaussovy jádrové funkce popsané v rovnici 9. [12]

$$K_{Gauss}(x) = (2\pi)^{-d/2} \cdot \exp\left(-\frac{1}{2}x^T x\right), \quad (9)$$

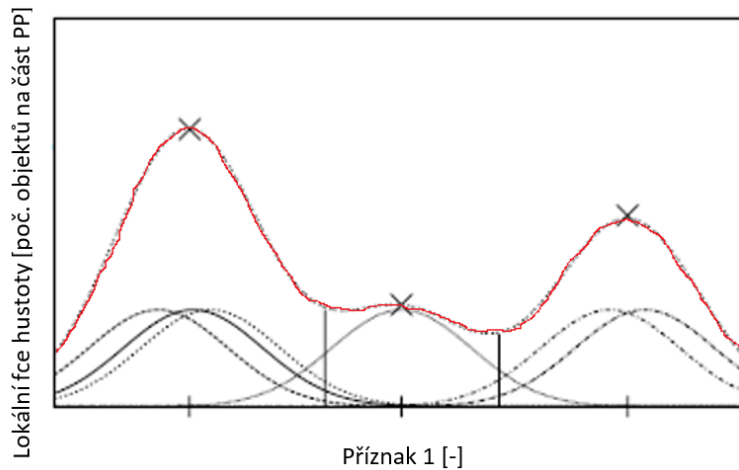
kde $K_{Gauss}(x)$ je funkce Gaussova jádra pro KDE v objektu x , d je dimenze (počet příznaků) dat a x je počítaný objekt [12].



Obrázek 6: Příklady jádrových funkcí používaných v KDE. Na obrázku vlevo je vidět čtvercová jádrová funkce a na obrázku vpravo Gaussova jádrová funkce. Na ose y je lokální funkce hustoty (jednotkou je počet objektů na část příznakového prostoru) daného jádra a osa x představuje vzdálenost od počítaného bodu umístěného ve středu jádrové funkce. Převzato z [12], upraveno.

Pro každý objekt je pomocí KDE počítána lokální funkce hustoty daného objektu (viz rovnice 10). Součtem lokálních funkcí hustoty všech objektů v prostoru získávám celkovou funkci hustoty pro daný příznakový prostor. Na obrázku 7 lze vidět příklad takového skládání lokálních funkcí hustoty pro 1-dimenzionální (1D) data. [12]

Příklad vzniku celkové fce hustoty v 1D příznakovém prostoru



Obrázek 7: Princip skládání celkové funkce hustoty (červená čára označená křížky) z lokálních funkcí hustoty pro jednotlivé objekty příznakového prostoru (ostatní čáry, kdy každá čára značí lokální funkci hustoty pro jeden objekt) v 1D prostoru. Osa x značí bezrozměrný příznakový prostor a osa y určuje funkci hustoty v počtu objektů na část příznakového prostoru. Převzato z [12], upraveno.

$$f^D(x) = \frac{1}{Nh^d} \sum_{i=1}^N K\left(\frac{1}{h}(x - x_i)\right), \quad (10)$$

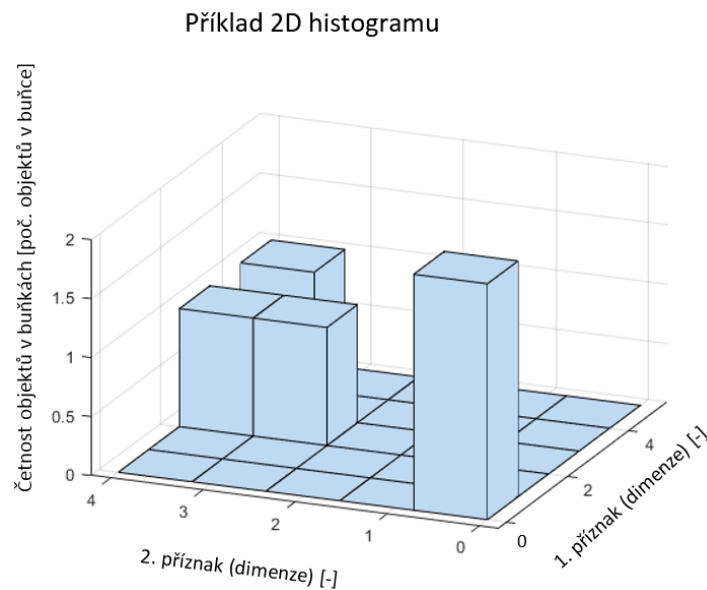
kde $f^D(x)$ je lokální funkce hustoty objektu x , N je počet objektů v počítaném prostoru, h je parametr hladkosti (viz strana 34), K je jádrová funkce KDE, x je objekt, pro nějž počítáme funkci hustoty a x_i je i -tý objekt.

Pokud do rovnice 10 dosadíme vztah z rovnice 9, získáváme vztah pro výpočet lokální funkce hustoty KDE za pomoci Gaussova jádra, jak lze vidět v rovnici 11.

$$f_{Gauss}^D(x) = \frac{1}{Nh^d (2\pi)^{d/2}} \sum_{i=1}^N \exp\left(-\frac{1}{2h^2}(x - x_i)^T(x - x_i)\right), \quad (11)$$

kde $f_{Gauss}^D(x)$ je lokální funkce hustoty objektu x pro Gaussovo jádro, N je počet objektů v počítaném prostoru, h je parametr hladkosti (viz strana 34), x je objekt, pro nějž počítáme funkci hustoty a x_i je i -tý objekt.

KDE je relativně složitý, proto algoritmus DENCLUE optimalizuje tento matematický základ [12]. Pro vytvoření celkové funkce hustoty dat se využívá multidimenzionální histogram namísto výpočtu KDE. Multidimenzionální histogram, stejně jako klasický histogram, počítá četnost objektů v jednotlivých buňkách. Tímto výpočtem získává hustotu objektů v buňkách. Na rozdíl od klasického histogramu se zde využívají buňky v tolika dimenzích, kolik je využívaných příznaků (v našem případě až 24 dimenzí). V každé dimenzi je pro zjednodušení rozdělen příznakový prostor na stejný počet buněk. Na obrázku 8 lze vidět příklad vytvořeného histogramu ve 2 dimenzích (2D). [12]

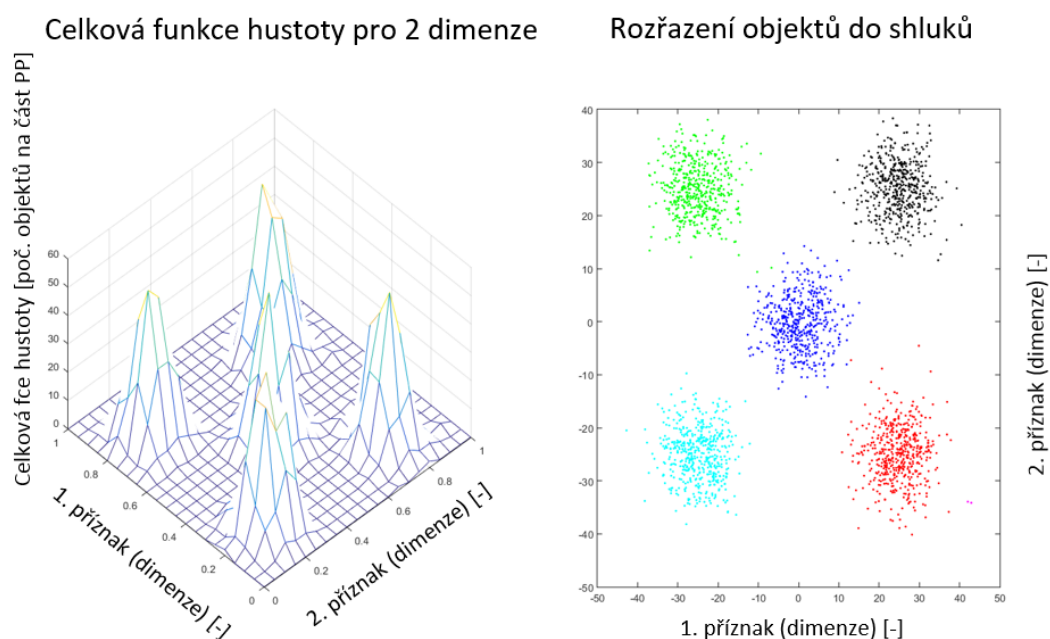


Obrázek 8: Příklad histogramu ve 2 dimenzích pro objekty rozdělené do 5 buněk v každé dimenzi.

Problémem vícerozměrného histogramu je nepravděpodobnost vysoké obsazenosti buněk u velkého počtu dimenzí [12]. Tento problém zlepšuje použití ASH (Average Shifted Histograms - průměr posunutých histogramů). ASH využívá citlivosti histogramu na počáteční místo histogramu. Mění u stejných dat volbu počátku histogramu a z jednotlivých histogramů získává zprůměrováním jeden výsledný histogram. V algoritmu DENCLUE se využívá $(d + 2)$ posunutí, kdy d je počet dimenzí (příznaků) počítaných objektů. Počet posunutí je dán posunutím vždy jedné dimenze, zatímco ostatní dimenze zůstávají bez posunutí. Do průměrování histogramů se dále započítává neposunutý multidimenzionální histogram a multidimenzionální histogram celý posunutý. Posunutí se děje

vždy o konstantní a malý krok. Pokud je krok posunutí nekonečně malý, blíží se výsledná funkce hustoty hustotě lokálního jádra. [12]

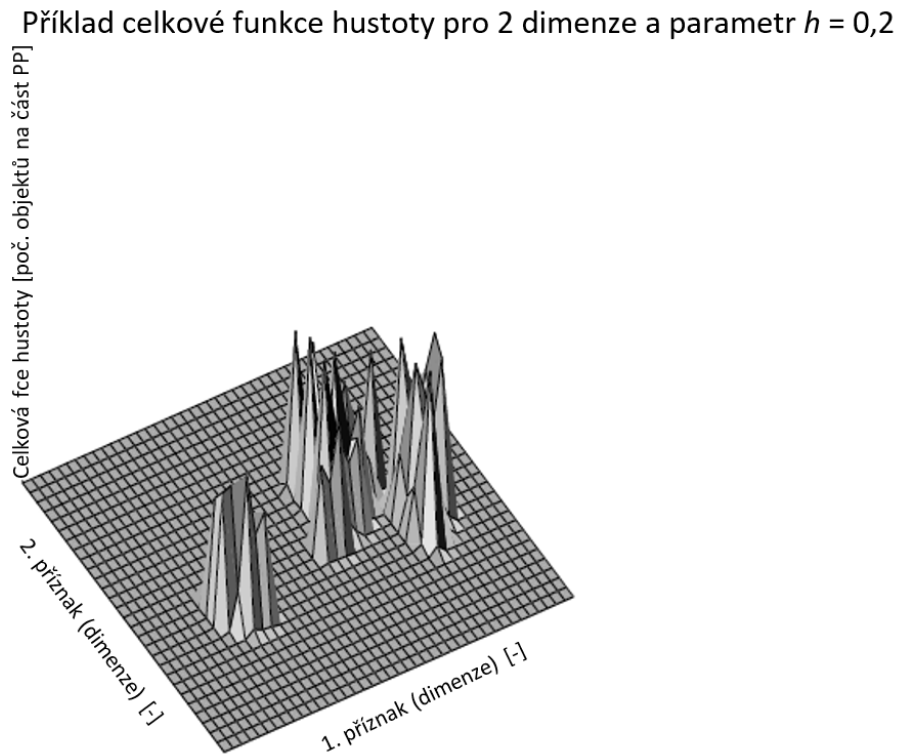
Po vytvoření celkové funkce hustoty z jednotlivých objektů je nutné rozdělit objekty do shluků. Toto rozdělení je provedeno za pomoci gradientu. Gradientem se určí lokální maxima celkové funkce hustoty. Ty reprezentují centra jednotlivých shluků. Výsledkem je stejný počet shluků, jako je počet lokálních maxim celkové funkce hustoty. Objekty, které spadají pod jedno lokální maximum celkové funkce hustoty, patří do stejného shluku jako dané lokální maximum. Princip zařazení objektů do shluků lze vidět na obrázku 9. [12]



Obrázek 9: Vpravo lze vidět příklad objektů ve 2D příznakovém prostoru, které jsou rozděleny do shluků (každý shluk reprezentují objekty totožné barvy) podle celkové funkce hustoty vytvořené z těchto objektů. Celková funkce hustoty je ukázána na obrázku vlevo. Jednotkou celkové funkce hustoty je počet objektů na část příznakového prostoru.

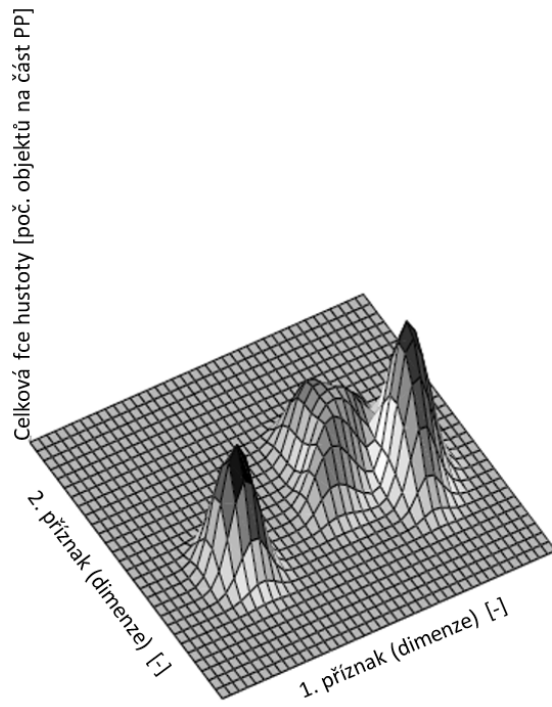
Algoritmus DENCLUE využívá dva parametry. Prvním parametrem je koeficient hladkosti h a druhým parametrem je koeficient šumu ξ . Význam koeficientu hladkosti je patrný z rovnice 11 na straně 31. Koeficient hladkosti udává jak moc hustotu počítaného objektu ovlivňují objekty z jeho okolí. Čím větší je h , tím více objektů z širšího okolí ovlivňuje lokální funkci hustoty. Zároveň tím „hladší“ je celková funkce hustoty a menší počet lokálních minim celkové funkce hustoty (počet shluků). Tento princip lze vidět

na obrázcích 10 a 11. V mém kódu nenastavuje uživatel koeficient hladkosti, ale nastavuje počet buněk histogramu v jedné dimenzi k . Tato změna umožňuje uživateli algoritmu lepší představu o významu parametru. To vede ke snazší orientaci i nepřiliš znalých uživatelů a tím zajištění správné funkčnosti kódu. Počet buněk histogramu svým významem odpovídá převrácené hodnotě koeficientu hladkosti, jak lze vidět na obrázcích 10 až 13. Čím méně bude buněk histogramu, tím více objektů z většího okolí ovlivňuje lokální hustotu. Zároveň tím „hladší“ je celková funkce hustoty. [12, 13]

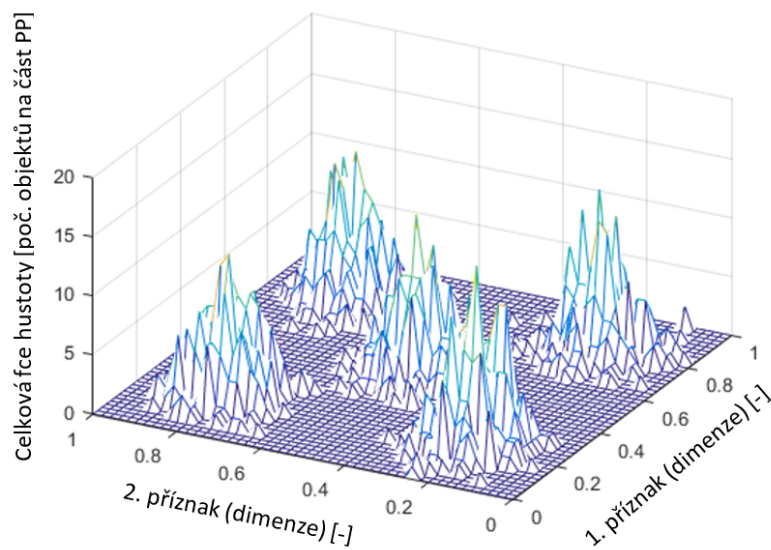


Obrázek 10: Příklad celkové funkce hustoty vytvořené autory algoritmu DENCLUE pro koeficient hladkosti $h = 0,2$. Na ose x a y jsou jednotlivé příznaky (bezrozměrné) a na ose z celková funkce hustoty mající jednotku počet objektů na část příznakového prostoru. Převzato z [12], upraveno.

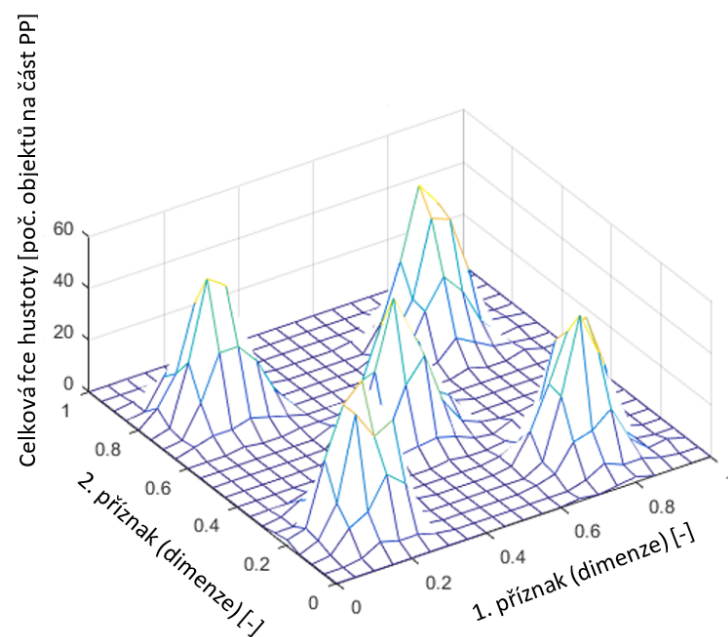
Příklad celkové funkce hustoty pro 2 dimenze a parametr $h = 0,6$



Obrázek 11: Příklad celkové funkce hustoty vytvořené autory algoritmu DENCLUE pro koeficient hladkosti $h = 0,6$. Na ose x a y jsou jednotlivé příznaky (bezrozměrné) a na ose z celková funkce hustoty mající jednotku počet objektů na část příznakového prostoru. Převzato z [12], upraveno.

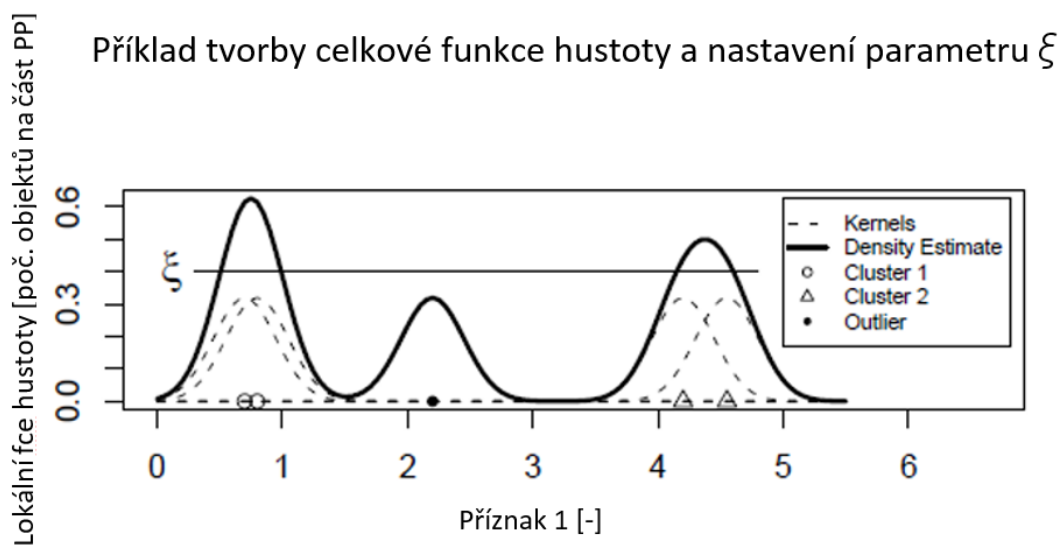
Příklad celkové funkce hustoty pro 2 dimenze a parametr $k = 50$ 

Obrázek 12: Příklad celkové funkce hustoty vytvořené pro počet buněk histogramu $k = 50$. Na ose x a y jsou jednotlivé příznaky (bezrozměrné) a na ose z je celková funkce hustoty mající jednotku počet objektů na část příznakového prostoru.

Příklad celkové funkce hustoty pro 2 dimenze a parametr $k = 20$ 

Obrázek 13: Příklad celkové funkce hustoty vytvořené pro počet buněk histogramu $k = 20$. Na ose x a y jsou jednotlivé příznaky (bezrozměrné) a na ose z je celková funkce hustoty mající jednotku počet objektů na část příznakového prostoru.

Koeficient šumu se vztahuje k lokálnímu maximu celkové funkce hustoty. Pokud je lokální maximum této funkce vyšší než ξ , lze lokální maximum považovat za centrum samostatného shluku. V opačné případě jsou všechny objekty patřící pod toto lokální maximum považovány za šum (viz obrázek 14). Pod pojmem šum si lze představit ne-relevantní objekty dat [12]. Mezi ně patří uniformně rozvržené objekty po příznakovém prostoru a vzdálené objekty. Vzdálené objekty jsou umístěné samostatně v příznakovém prostoru. Pokud by byl koeficient šumu rovný nule, mohl by každý jeden objekt být samostatným shlukem. Správné nastavení koeficientu šumu zajišťuje zaměření algoritmu pouze na hustotně zaplněné oblasti a snižuje časovou náročnost algoritmu. [12]



Obrázek 14: Princip funkce parametru ξ , kde kolečka a trojúhelníky (Cluster 1, Cluster 2, Outlier) znamenají pozici objektů v datovém prostoru, čárkované čáry představují lokální funkce hustoty (Kernels), silná spojitá čára značí celkovou funkci hustoty (Density Estimate) a slabší rovnoběžná čára znamená úroveň koeficientu ξ . Osa x určuje bezrozměrný příznakový prostor a osa y určuje velikost funkce hustoty. Z obrázku je vidět, že prázdná kolečka a trojúhelníky (Cluster 1 a Cluster 2) vytváří samostatné shluky a plné kolečko (Outlier) je zařazeno do šumu. Převzato z [13], upraveno.

Pseudokód 1 ukazuje princip algoritmu DENCLUE používaného v této práci.

Vstupy: objekty v příznakovém prostoru, k , ξ

Výstupy: rozřazené objekty do shluků

načtení objektů;

for $i = 1 : \text{počet příznaků} + 2$ **do**

 rozdělení příznaků na k buněk histogramu;

 provedení i -tého posunutí;

 rozřazení objektů do buněk histogramu;

end

zprůměrování posunutých histogramů;

zjištění lokálních maxim ASH;

for $j = 1 : \text{počet lok. maxim}$ **do**

if *amplituda j -tého lok. maxima* $> \xi$ **then**

 vytvoření centra nového shluku z j -tého lok. maxima;

 přiřazení objektů patřících pod j -té lok. maximum do stejného shluku;

else

 zařazení j -tého lok. maxima do šumového shluku;

 přiřazení objektů patřící pod j -té lok. maximum do šumového shluku;

end

end

Pseudokód 1: Automatická klasifikace algoritmem DENCLUE používaným v této práci.

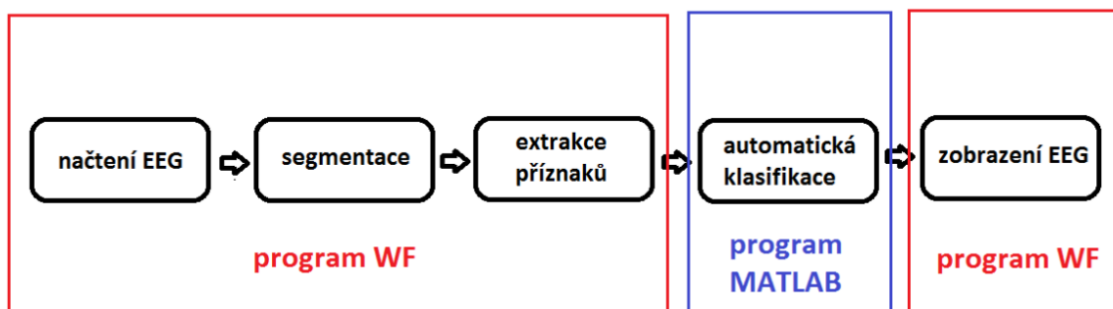
2 Metodologie

Před započítím samotného měření bylo potřeba určit nastavení používaných programů, aby výsledky jednotlivých automatických klasifikací signálu byly porovnatelné. Toto nastavení diskutuji v kapitole 2.1. V kapitole 2.1.2 popisují vytvoření grafického uživatelského rozhraní (GUI - Graphic User Interface). Pomocí GUI si může uživatel požadovaný algoritmus DENCLUE pro vybraná data snadno spustit a nastavit u něho potřebné parametry. Popis používaných dat zajistí, aby výsledky diplomové práce byly reprodukovatelné. Toto popisují v kapitole 2.2. Pro zajištění kvalitní a přesné interpretace výsledků bylo potřeba určit správnou metodu statistického vyhodnocení dat. Metodiku statistického vyhodnocení dat popisují v kapitole 2.3.

EEG záznam je velmi specifický již svojí stochastickou povahou [3]. To činí zvláštní nároky na používaný algoritmus automatické klasifikace. Bylo potřeba algoritmus DENCLUE upravit, aby správně klasifikoval EEG záznam podle podmínek uvedených v kapitole 2.3. Úpravu algoritmu jsem popsal v kapitole 2.4. Pro ověření vylepšení vlastností algoritmu jsem statisticky zpracoval i výsledek automatické klasifikace neupraveného algoritmu DENCLUE. Následně jsem porovnal výsledky upraveného a neupraveného algoritmu DENCLUE při aplikaci na kompletní EEG záznam (viz kapitoly 3.2.2 a 4.2).

2.1 Programová specifikace

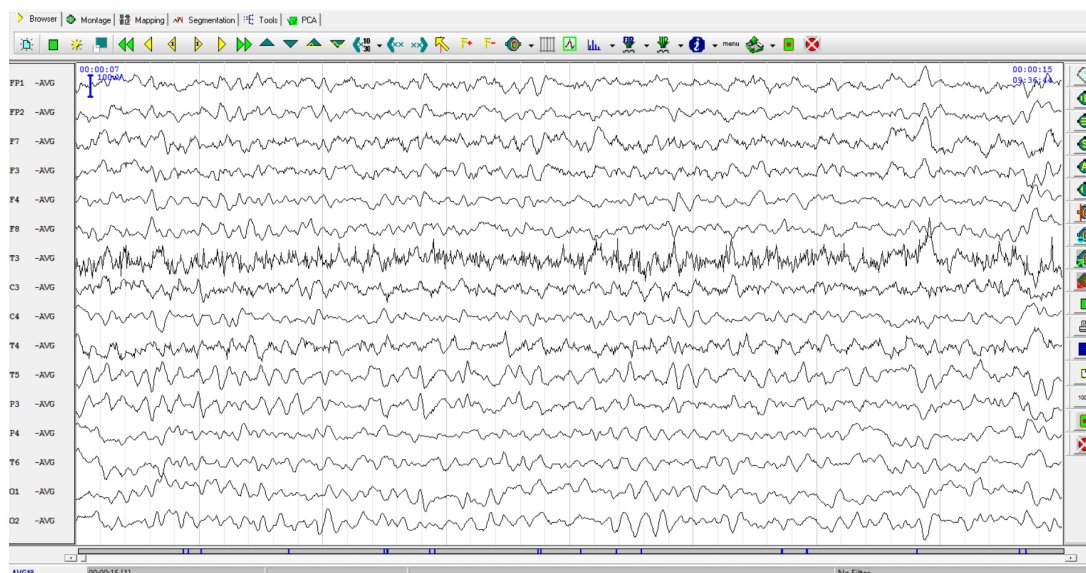
V této práci jsem využíval program WF [5] a program MATLAB R2015a [26]. Na obrázku 15 lze vidět blokové schéma znázorňující funkci jednotlivých programů u automatické klasifikace EEG záznamu. Blokové schéma znázorňuje automatickou klasifikaci EEG záznamu algoritmem DENCLUE.



Obrázek 15: Blokové schéma znázorňující automatickou klasifikaci EEG záznamu v této práci. Princip blokového schématu byl využit pro automatickou klasifikaci algoritmem DENCLUE. Pro automatickou klasifikaci byly využity programy WF a MATLAB R2015a [5, 26].

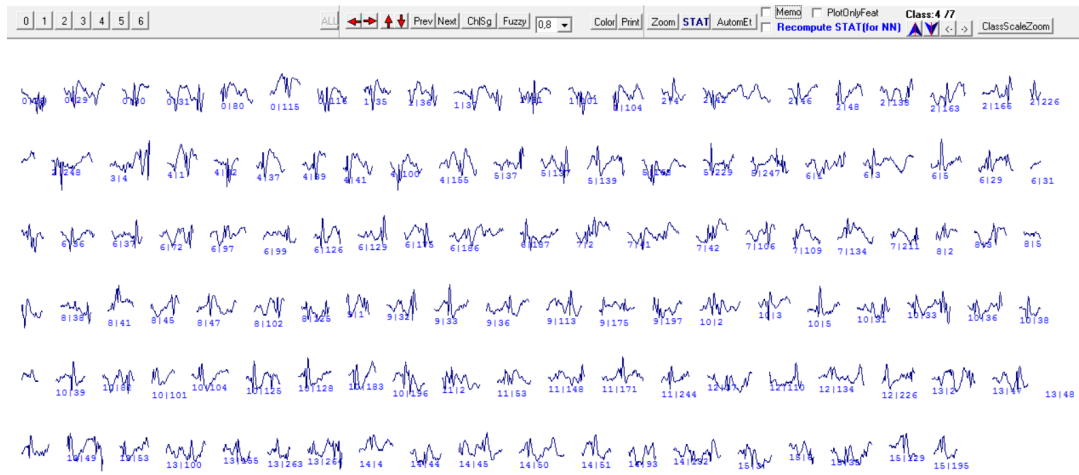
2.1.1 Wafe-Finder (WF)

Pro načtení signálu EEG, jeho segmentaci, extrakci příznaků a vykreslení jsem používal program WF. Program WF je v klinické praxi používaný program pro načtení a zpracování EEG signálu. Program WF jsem využíval i k vizualizaci výsledků automatické klasifikace všech používaných algoritmů. Dále jsem porovnával vytvořené algoritmy automatické klasifikace s automatickou klasifikací k-means. Ta je součástí programu WF a využívá se v klinické praxi. Díky tomuto porovnání jsem ověřil praktické využití naprogramovaných algoritmů. Na obrázku 16 je zobrazeno prostředí programu WF při načtení EEG signálu. [5]



Obrázek 16: Příklad prostředí v praxi používaného programu WF, využívaného v této práci pro vizualizaci a předzpracování EEG dat [5].

Program WF načítá naměřené EEG záznamy. Po provedení adaptivní segmentace a extrakce příznaků vytváří program WF formát .nra. Formát .nra obsahuje normalizované příznaky pro každý segment záznamu. Program WF formát .nra načítá pro provedení automatické klasifikace. Stejný formát jsem v této práci využíval pro načtení příznakového prostoru. Díky využití programu WF odpadla nutnost provést segmentaci a následnou extrakci příznaků. Výstup automatické klasifikace byl uložen v dokumentu bitového formátu .qub. Tento formát načítá program WF pro vizualizaci výsledků automatické klasifikace (viz obrázek 17). [5]



Obrázek 17: Příklad zobrazení jednoho shluku vytvořeného automatickou klasifikací EEG záznamu. Číslo značí kanál daného segmentu a pořadí segmentu v tomto kanále. Příklad je zobrazen v prostředí programu WF, jenž je využíván v této práci pro vizualizaci a předzpracování EEG dat [5].

Program WF využívá adaptivní segmentaci EEG záznamu (viz kapitola 1.2). Pro tuto práci jsem využíval nastavení adaptivní segmentace podle doporučení doc. Ing. Vladimíra Krajčiči, CSc. Stejné nastavení se využívá u programu WF v klinické praxi. Konkrétní nastavené hodnoty pro adaptivní segmentaci lze vidět v tabulce 2.

Tabulka 2: Zkratka nastavovaného parametru adaptivní segmentace v programu WF, její český popis a hodnota parametru, kterou jsem využíval v této práci pro vzorkovací frekvenci 128 Hz po doporučení doc. Ing. Vladimíra Krajčiči, CSc.

Zkratka par.	Popis parametru	Hodnota par.
Window Length	délka dvojitého okna	128 vzorků
G Window Length	délka vnitřního okna	15 vzorků
STEP	krok	1 vzorků
Optim	optimalizace	1 vzorků
MINLENGTH	minimální délka segmentu	64 vzorků
Number of Scan pts	posunutí pro nalezení hrany	15 vzorků
Trashold	hranice pro začátek segmentu	81 [-]
Max Segm Length	maximální délka segmentu	1024 vzorků

Program WF může využívat extrakci 24 příznaků, které jsou popsány v kapitole 1.2.1. Vliv extrakce různých příznaků na porovnávání algoritmů bylo potřeba co nejvíce snížit.

Z tohoto důvodu jsem pro neupravený algoritmus DENCLUE používal stejné příznaky, jako využívá program WF pro algoritmu k-means. Pro algoritmus DENCLUE uzpůsobenou na EEG záznam jsem musel změnit používané příznaky (viz kapitola 2.4). Stále jsem ovšem vybíral příznaky ze sady příznaků počítaných programem WF. [5]

2.1.2 Modul v MATLABu

Jednotlivé algoritmy jsem programoval v programovacím prostředí MATLAB (Matrix Laboratory) R2015a [26]. V tabulce 3 jsou popsány základní funkce, které jsem v tomto programu vytvořil.

Tabulka 3: Název mnou vytvořených funkcí v programu MATLAB R2015a a jejich popis.

Název funkce	Popis funkce
start_DENCLUE.fig	otevření GUI a spuštění DENCLUE
start_DENCLUE.m	otevření GUI a spuštění DENCLUE
downl.m	načtení, upravení a výpis EEG dat
DENCLUE.m	provedení algoritmu DENCLUE
ASH.m	provedení ASH
MDHist.m	provedení multidimenzionálního histogramu
AvailebHist.m	určuje zaplněnost buněk napříč posunutím
FindNeigCell.m	hledá sousední buňky v multidimenzionálním prostoru
FindRows.m	hledá sloupce stejných prvků z matice a vektoru pro ASH
FindRowsMD.m	hledá sloupce stejných prvků z matice a vektoru pro MDHist

V prostředí MATLAB R2015a jsem vytvořil jednoduché GUI. GUI využívá uživatel mého kódu ke spuštění automatické klasifikace EEG záznamu a lze ho vidět na obrázku 18. V GUI si uživatel vybere, zda chce použít algoritmus uzpůsobený na EEG záznam či neuzpůsobený. Poté se uživateli otevře okno, ve kterém uživatel vybere soubor formátu .nra. Poté algoritmus provede svojí činnost. Výsledek bude uložen ve formě souboru formátu .qub. Uložen bude do stejné složky, z které uživatel načítal soubor formátu .nra.

Celý algoritmus bude v nejbližší době aplikován do GUI programu, který vytváří celý EEG tým pro načtení a zpracování EEG záznamu.

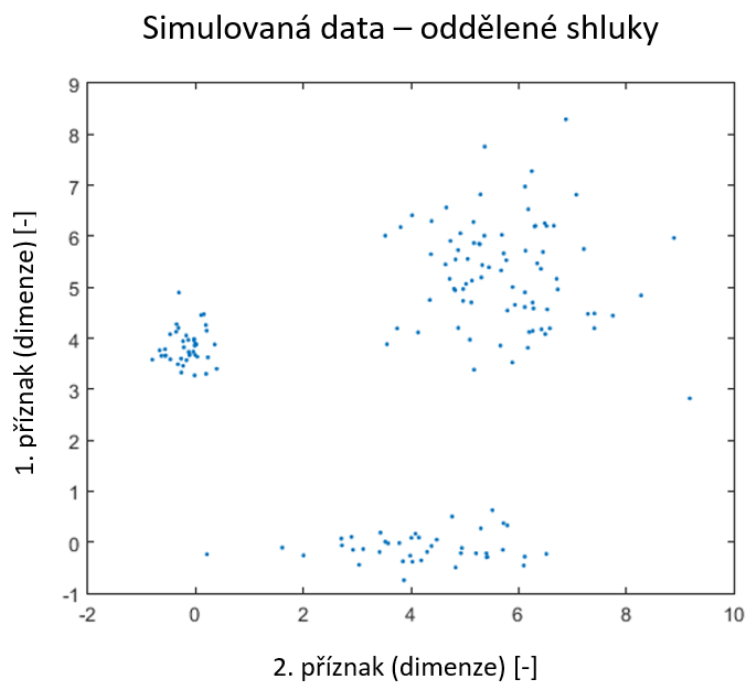


Obrázek 18: Jednoduché GUI využívané uživateli pro spuštění algoritmu DENCLUE. Uživatel může spustit na EEG záznam upravenou nebo neupravenou verzi algoritmu DENCLUE.

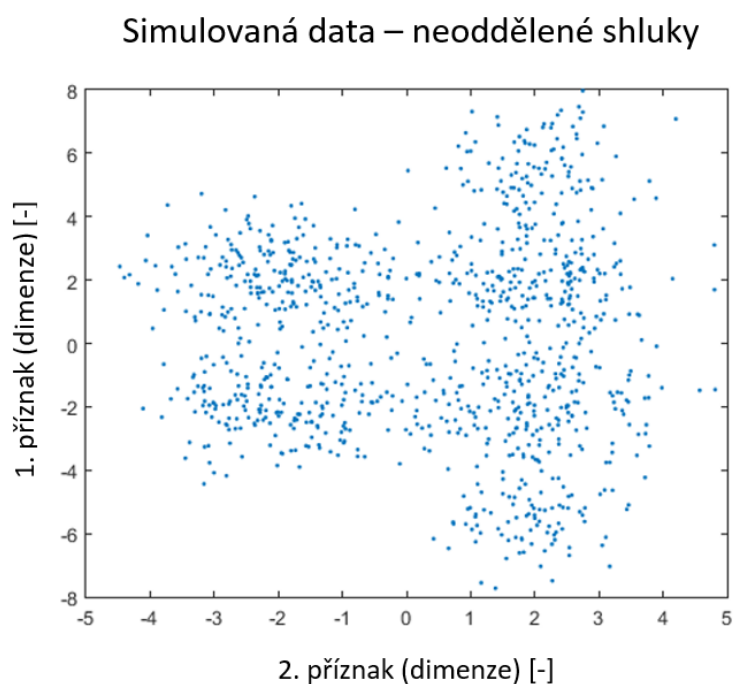
2.2 Použitá data

2.2.1 Simulovaná data

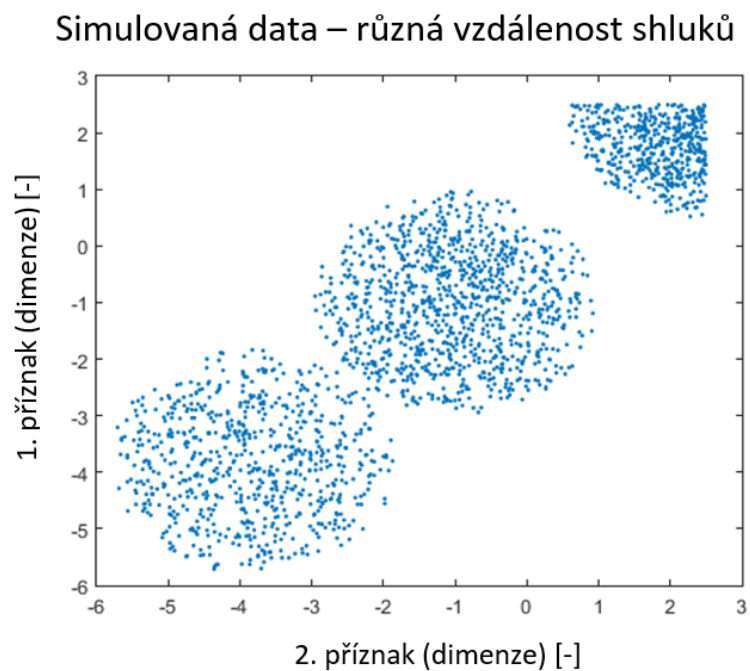
Pro zjištění použitelnosti jednotlivých algoritmů jsem je nejprve vyzkoušel pro simulovaná data. Simulovaná data jsem vytvořil v programu MATLAB R2015a [26]. Vytvořil jsem 4 sady simulovaných dat. Jednalo se o výrazně oddělené shluky (viz obrázek 19), nevýrazně oddělené shluky (viz obrázek 20), shluky s různou vzájemnou vzdáleností (viz obrázek 21) a prolnuté shluky (viz obrázek 22). Jejich statistické vyhodnocení lze pozorovat v kapitolách 3.1 a 4.1. Na trénovací data jsem aplikoval algoritmus k-means a DENCLUE neupravený na EEG data. Algoritmus DENCLUE upravený na EEG data jsem testoval pouze na EEG záznamech.



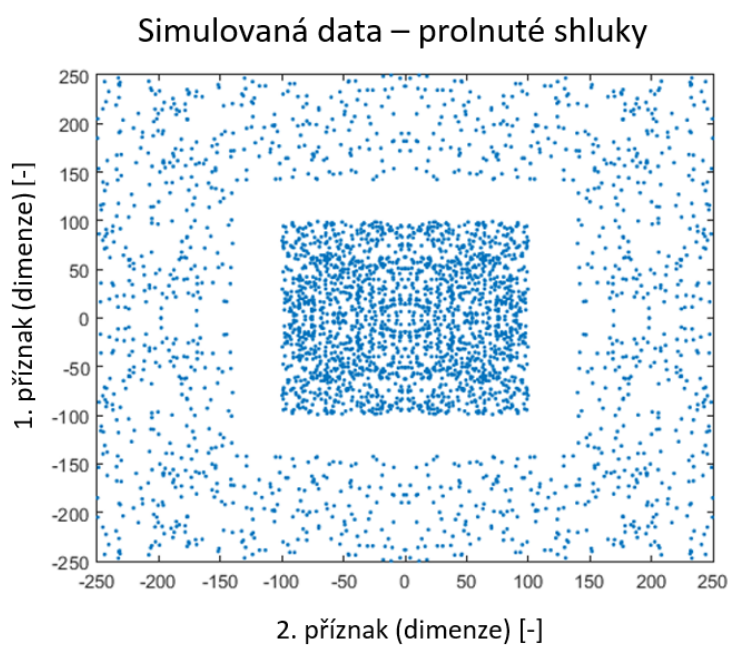
Obrázek 19: Simulovaná data vytvořená pro testování algoritmů automatické klasifikace. Data simulují zřetelně oddělené shluky



Obrázek 20: Simulovaná data vytvořená pro testování algoritmů automatické klasifikace. Data simulují nevýrazně oddělené shluky



Obrázek 21: Simulovaná data vytvořená pro testování algoritmů automatické klasifikace. Data simulují shluky s různou vzájemnou vzdáleností



Obrázek 22: Simulovaná data vytvořená pro testování algoritmů automatické klasifikace. Data simulují prolnuté

2.2.2 Reálné EEG záznamy

Testované a porovnávací algoritmy jsem otestoval na reálně naměřeném záznamu EEG. V této práci jsem používal data naměřená v Nemocnici Na Bulovce přístrojem BRA-INQUICK. Data byla měřena během klinického vyšetření na subjektech s podezřením na epilepsii. Všechna měření schválila etická komise Nemocnice Na Bulovce v roce 2011. Každý subjekt před měřením podepsal informovaný souhlas. Věk subjektů se pohyboval mezi 26 až 60 roky. Jednalo se o subjekty různého pohlaví.

Pro tuto pilotní studii jsem využíval EEG záznamy z 15 subjektů. Záznamy byly v zájmu zachování etiky anonymizovány. Všechny záznamy byly z multikanálového měření. Konkrétně se jednalo o záznamy z šestnácti kanálového EEG.

Pro statistické zpracování EEG dat byl podstatný počet segmentů v každém záznamu. Pro jednotlivé segmenty expert určoval jejich správné zařazení do shluků. Počet segmentů EEG záznamu závisí na délce měření, počtu kanálů a členitosti EEG záznamu. Počet segmentů se u dodaných EEG záznamů pohyboval mezi 8 957 a 42 038 segmenty na záznam.

2.3 Statistické vyhodnocení dat

2.3.1 Kvantitativní analýza

Během kvantitativního statistického vyhodnocení simulovaných dat jsem provedl vyhodnocení časové náročnosti algoritmu DENCLUE a algoritmu k-means. Vyhodnocení proběhlo v závislosti na počtu vyhodnocovaných segmentů. U reálného EEG záznamu jsem kvantitativní analýzu prováděl také v závislosti na počtu vyhodnocovaných segmentů. Jednalo se opět o vyhodnocení časové náročnosti algoritmů. Kvantitativní analýzu reálného EEG záznamu jsem prováděl pro algoritmus k-means, DENCLUE neupravený na EEG záznam a DENCLUE upravený na EEG záznam (viz kapitola 2.4).

2.3.2 Kvalitativní analýza

Kvalitativní část statistického vyhodnocení se zabývala správností zařazení jednotlivých segmentů do shluků. Pro tyto účely jsem použil ROC analýzu. ROC (Receiver Operating Characteristics) analýza je dlouhodobě užívána v teorii detekce signálu [27]. Pomocí

ROC analýzy se určuje přesnost zařazení jednotlivých elementů do jedné ze dvou skupin (pozitivní skupina a negativní skupina). Během ROC analýzy se každému elementu přiřadí jedno z následujících označení. Označení vyjadřují správnost zařazení elementu do skupin. Jedná se o označení TP (true positive - správně pozitivní), TN (true negative - správně negativní), FP (false positive - špatně pozitivní) a FN (false negative - špatně negativní). Z těchto označení elementů se počítá senzitivita a specificita skupiny. Pomocí rovnice 12 se vypočítává senzitivita. Senzitivita vyjadřuje počet elementů správně zařazených mezi pozitivní skupinu vůči všem elementům, které by měly být zařazeny mezi pozitivní skupinu. Rovnice 13 ukazuje výpočet specificity. Specificita ukazuje počet elementů správně zařazených do negativní skupiny vůči všem elementům, které by měly patřit do negativní skupiny. [27, 28, 29]

$$\text{senzitivita} = \frac{TP}{TP + FN}, \quad (12)$$

kde *senzitivita* je počítaná senzitivita, *TP* je počet elementů správně zařazených do pozitivní skupiny a *FN* je počet elementů špatně zařazených do negativní skupiny [28].

$$\text{specificita} = \frac{TN}{TN + FP}, \quad (13)$$

kde *specificita* je počítaná specificita, *TN* je počet elementů správně zařazených do negativní skupiny a *FP* je počet elementů špatně zařazených do pozitivní skupiny [28].

Dalším významným parametrem počítaným v ROC analýze je prediktivní hodnota pozitivního testu (PPV). PPV se počítá podle vzorce uvedeného v rovnici 14. Pomocí PPV se zjišťuje pravděpodobnost, že element vybraný do pozitivní skupiny, do této skupiny opravdu patří. [29]

$$PPV = \frac{TP}{TP + FP}, \quad (14)$$

kde *PPV* je prediktivní hodnota pozitivního testu, *TP* je počet elementů správně zařazených do pozitivní skupiny a *FP* je počet elementů špatně zařazených do pozitivní

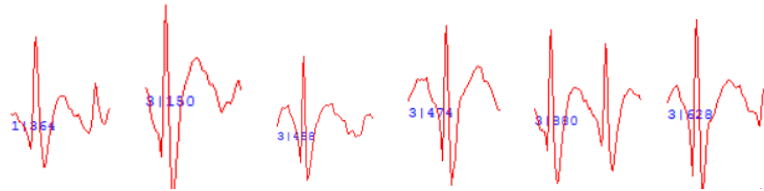
skupiny [29].

V této práci jsem za pomoci ROC analýzy určoval správnost zařazení segmentů EEG záznamu a simulovaných dat do shluků. Senzitivitu, specificitu a PPV jsem počítal pro jednotlivé shluky i pro celý EEG záznam, popřípadě simulovaná data. Označení segmentů se dělo vždy pro každý konkrétní shluk. Pokud byl segment zařazen algoritmem automatické klasifikace správně do tohoto shluku, byl tento segment označen jako TP. Pokud byl v tomto shluku zařazen segment, který do něj nepatřil, byl označen FP. Pokud existoval segment patřící do tohoto shluku, ale byl zařazen do jiného shluku, byl označen jako FN. Ostatní segmenty byly pro tento shluk zařazeny jako TN.

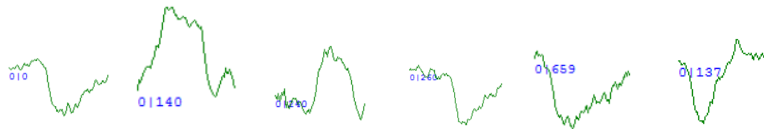
V každém záznamu EEG jsem rozlišoval 6 základních shluků. Tyto shluky byly vybírány na základě významu jednotlivých segmentů. Při jejich výběru jsem vycházel z práce A. T. Tzallase [30]. V této práci se rozlišoval shluk fyziologické aktivity mozku, shluk epileptických grafoelementů, shluk EMG artefaktů a shluk pomalých očních artefaktů. Já jsem rozlišoval kromě zmíněných shluků i shluky impulzních artefaktů (výrazný pozitivní hrot - hrot směřující dolů) a rovných čar (část záznamu EEG, kdy nabyl snímán EEG signál). Tyto dvě části EEG záznamu mají výrazně odlišný charakter a při klasifikaci podle práce [30] by nebylo kam je zařadit. Zároveň jsem v této práci do shluku EMG artefaktů přiděloval segmenty výrazně zašuměné artefaktem síťový brum. Mezi oběma typy artefaktů je značná podoba. Na obrázcích 23 až 28 jsou vidět příklady typických segmentů pro shluky, rozlišované v této práci.



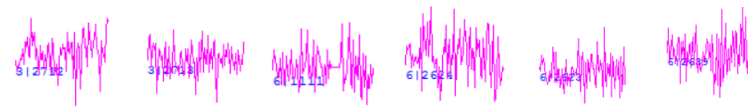
Obrázek 23: Příklad segmentů typicky zastupujících shluk fyziologické aktivity mozku. Číslice u každého segmentu značí číslo kanálu a pořadí segmentu v kanálu. Segmenty jsou extrahované z reálného záznamu EEG zobrazeného programem WF [5].



Obrázek 24: Příklad segmentů typicky zastupujících shluk epileptických grafoelementů. Číslice u každého segmentu značí číslo kanálu a pořadí segmentu v kanálu. Segmenty jsou extrahované z reálného záznamu EEG zobrazeného programem WF [5].



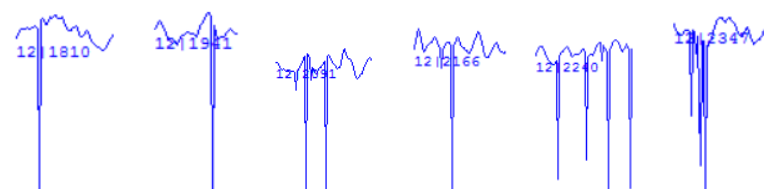
Obrázek 25: Příklad segmentů typicky zastupujících shluk pomalých očních artefaktů. Číslice u každého segmentu značí číslo kanálu a pořadí segmentu v kanálu. Segmenty jsou extrahované z reálného záznamu EEG zobrazeného programem WF [5].



Obrázek 26: Příklad segmentů typicky zastupujících shluk EMG artefaktů a zašuměného signálu. Číslice u každého segmentu značí číslo kanálu a pořadí segmentu v kanálu. Segmenty jsou extrahované z reálného záznamu EEG zobrazeného programem WF [5].



Obrázek 27: Příklad segmentů typicky zastupujících shluk rovných čar. Číslice u každého segmentu značí číslo kanálu a pořadí segmentu v kanálu. Segmenty jsou extrahované z reálného záznamu EEG zobrazeného programem WF [5].



Obrázek 28: Příklad segmentů typicky zastupujících shluk impulzních artefaktů. Číslice u každého segmentu značí číslo kanálu a pořadí segmentu v kanálu. Segmenty jsou extrahované z reálného záznamu EEG zobrazeného programem WF [5].

Každý algoritmus vytváří různý počet shluků. Počet vytvořených shluků se mohl lišit i pro různé EEG záznamy. Bylo tedy potřeba nejprve u každého EEG záznamu určit, které z vytvořených shluků reprezentují 6 hledaných shluků. Existovaly zde 3 krajní případy. Hledaný shluk se v záznamu nevyskytoval, více vytvořených shluků reprezentovalo segmenty hledaného shluku a hledaný shluk byl v záznamu obsažen, ale jeho segmenty netvořily samostatný shluk. Prvním krajním případem byla situace, kdy se ve výsledném záznamu nevyskytoval ani jeden segment z hledaného shluku. Například v celém záznamu nebyla ani jedna rovná čára. V takovém případě pro něj nebyla počítána senzitivita, specificita a PPV. Druhým krajním případem byla situace, kdy algoritmus vytvořil dva a více shluků, v kterých dominovaly stejné segmenty. V takovém případě se vybral za hledaný shluk ten, který obsahoval více těchto segmentů. Stejně segmenty obsažené v ostatních shlucích byly brány jako FN. Třetí krajní situace vznikala, pokud v záznamu existovaly segmenty hledaného shluku, ale ani o jednom vytvořeném shluku se nedalo říci, že tyto segmenty reprezentuje. V tomto případě bylo zapsáno, že algoritmus nenalezl hledaný shluk.

Zařazení do správného shluku určovali dva experti se zkušenostmi z klinické praxe. Jednalo se o doc. Ing. Vladimíra Krajčů, CSc. a prim. MUDr. Ing. Svojmila Petránka, MBA. Každý expert dostal všechny segmenty rozdělené do shluků. Poté každý expert určil, zda jsou jednotlivé segmenty zařazeny správně. V případě, že segmenty nebyly zařazeny správně, určil expert do jakých shluků by měly být dané segmenty zařazeny. Pokud se experti u některých segmentů neshodovali, došlo na společnou konzultaci. Během konzultace se určilo správné zařazení diskutabilních segmentů. Z této analýzy byly předem odstraněny veškeré segmenty vzniklé špatnou segmentací. Příklady takovýchto segmentů jsou vidět na obrázku 29.



Obrázek 29: Příklady špatně nasegmentovaných segmentů, které jsou odstraněny ze statistického zpracování. Segmenty jsou vzaty z reálného záznamu EEG zobrazeného programem WF [5].

Statisticky jsem zpracovával EEG záznamy z 15 subjektů. Z nich jsem po adaptivní segmentaci získal celkem 317 920 segmentů. Kvalitativní vyhodnocení jsem rozdělil na dvě části. V první části jsem statisticky vyhodnotil kompletní 3 EEG záznamy. V druhé části jsem statisticky vyhodnotil 600 náhodných segmentů z 12 subjektů. Tři kompletní EEG záznamy měly dohromady 102 050 segmentů. Výsledek jejich statistického zpracování ukazuje, jak moc přesné jsou jednotlivé algoritmy při kompletní klasifikaci EEG záznamu. Statistické vyhodnocení této části jsem provedl zvlášť pro algoritmy k-means, DENCLUE neupravený na EEG záznam a DENCLUE upravený na EEG záznam (viz kapitola 2.4).

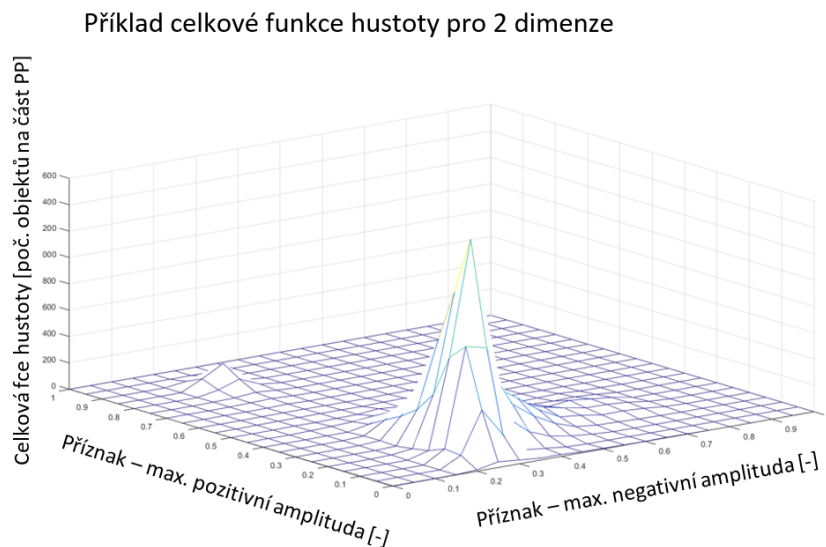
V druhé části kvalitativního hodnocení jsem pro 12 subjektů provedl náhodný výběr 50 segmentů u každého EEG záznamu. Náhodný výběr jsem provedl za pomoci programového prostředí MATLAB R2015a [26]. Konkrétně jsem ho provedl za pomoci funkce *rand()*. Touto funkcí jsem nejprve získal číslo stránky rozklasifikovaných segmentů. Následně jsem stejnou funkcí získal pořadí segmentu na stránce. Takto jsem vybral 50 náhodných segmentů pro algoritmus k-means. Tyto segmenty jsem hodnotil i u ostatních algoritmů automatické klasifikace. Důvodem byla snaha minimalizovat chybu, jenž může vzniknout výběrem různých segmentů pro různé algoritmy. Výsledky druhé části kvalitativního vyhodnocení ukazují rozlišnost výsledků automatické klasifikace pro různé EEG záznamy. Vyhodnocování jsem prováděl pro algoritmus k-means a DENCLUE upravený na EEG záznam (viz kapitola 2.4). Algoritmus DENCLUE neupravený na EEG záznam již vyhodnocován nebyl. Pro potvrzení jeho zjevné nefunkčnosti na EEG záznam postačily výsledky z kvalitativního vyhodnocení celého záznamu (viz kapitoly 3.2.2 a 4.2).

2.4 Úprava algoritmu DENCLUE na EEG data

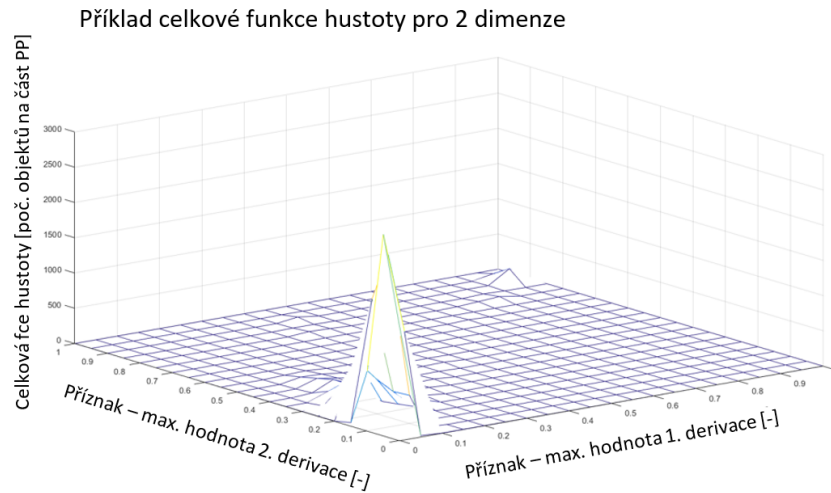
Algoritmus DENCLUE byl schopný u EEG rozlišit dva obsazené shluky a několik drobných shluků. Drobné shluky byly velmi málo zaplněné a navíc se od sebe málokdy odlišovaly. Tyto výsledky lze pozorovat v kapitole 3.2.2. Počet drobných shluků se lišil pouze nastavením parametrů (viz kapitola 2.4.1). Jeden více obsazený shluk obsahoval šumové segmenty. Druhý více obsazený shluk obsahoval všechny zbylé segmenty EEG záznamu. Bylo tedy viditelné, že tato metoda není vhodná pro aplikaci na EEG data. Toto potvr-

dily i pozdější výsledky (viz kapitoly 3.2.2 a 4.2). Z těchto důvodů bylo nutné algoritmus upravit.

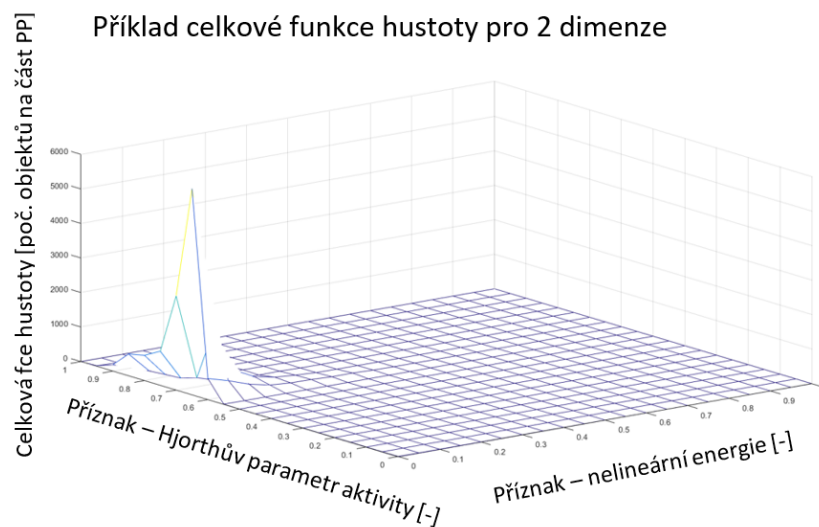
Algoritmus DENCLUE jsem upravil na základě specifčnosti EEG dat v příznakovém prostoru. Právě z toho důvodu jsem vizualizoval EEG data v příznakovém prostoru. To nebylo možné ve 23 dimenzích. Z toho důvodu jsem vykreslil veškeré kombinace 2 příznaků ze skupiny používaných příznaků. Na obrázcích 30 až 32 lze vidět příklady vykreslení 3 různých kombinací 2D příznakového prostoru pro EEG záznam. Na těchto obrázcích lze vidět, že všechny zobrazené příznaky měly totožnou charakteristiku. Veškerým datům v příznakovém prostoru hustotně vévodil 1 výrazný shluk. Tento charakter měly i všechny ostatní kombinace příznaků ve 2D příznakovém prostoru EEG.



Obrázek 30: Příklad celkové funkce hustoty. Na ose z je zobrazena celková funkce hustoty (mající jednotku počet objektů na část příznakového prostoru) pro 2 příznaky, vybrané z příznaků používaných programem WF [5]. Jedná se o příznaky „maximální pozitivní amplituda v daném segmentu“ a „maximální negativní amplituda v daném segmentu“. Na ose x a y se nachází bezrozměrný příznakový prostor.



Obrázek 31: Příklad celkové funkce hustoty. Na ose z je zobrazena celková funkce hustoty (mající jednotku počet objektů na část příznakového prostoru) pro 2 příznaky, vybrané z příznaků používaných programem WF [5]. Jedná se o příznaky „maximální hodnota 1. derivace v segmentu“ a „maximální hodnota 2. derivace v segmentu“. Na ose x a y se nachází bezrozměrný příznakový prostor.



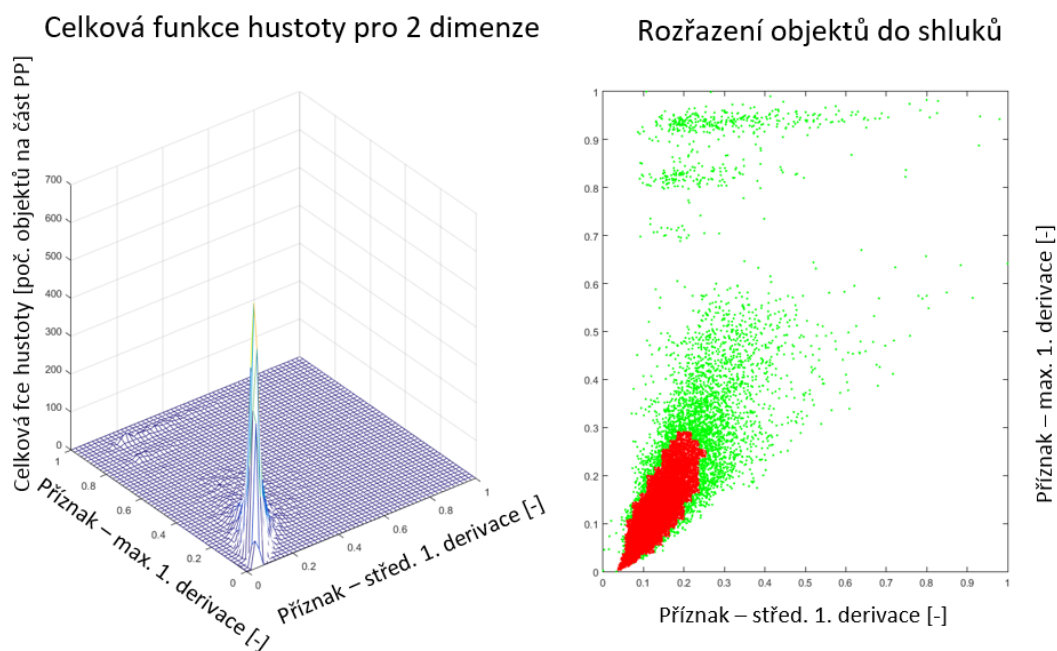
Obrázek 32: Příklad celkové funkce hustoty. Na ose z je zobrazena celková funkce hustoty (mající jednotku počet objektů na část příznakového prostoru) pro 2 příznaky, vybrané z příznaků používaných programem WF [5]. Jedná se o příznaky „Hjorthův parametr aktivity“ a „nelineární energie segmentu“. Na ose x a y se nachází bezrozměrný příznakový prostor.

Podobný charakter by měla mít i kombinace těchto příznaků ve dvaceti třech dimenzích příznakového prostoru. Velká amplituda celkové funkce hustoty jednoho shluku tedy mohla zastínit ostatní případné shluky, které byly oproti tomuto shluku nevýrazné.

Dosavadní zjištění ze zobrazení příznakového prostoru EEG dat jsem porovnal s charakterem EEG záznamu a s výsledky automatické klasifikace EEG neupraveným algoritmem DENCLUE (viz kapitoly 3.2.2 a 4.2). Výsledkem bylo zjištění, že většinový shluk představoval fyziologickou aktivitu mozku. Té je ve většině EEG záznamů nadbytek a celou ji zařazujeme do jednoho shluku (viz kapitola 2.3). V prvním kroku algoritmu jsem tedy shluk fyziologické aktivity mozku oddělil jako samostatný shluk. Ve druhém kroku algoritmu jsem pak klasifikoval zbytek záznamu samostatně. Ve druhém kroku se zvýraznily rozdíly mezi jednotlivými shluky neklasifikovanými do shluku fyziologické aktivity mozku. To vedlo k jejich snazšímu rozdělení (viz obrázek 34).

K oddělení shluku fyziologické aktivity mozku od ostatních shluků jsem použil oddělení šumu od hlavního shluku příznakového prostoru. Bylo nutné, aby šum obsahoval všechny shluky, které chceme klasifikovat odděleně od shluku fyziologické aktivity mozku (shluk epileptických grafoelementů, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu, impulzních artefaktů a shluk rovných čar - viz kapitola 2.3). Proto byl vybrán příznak první derivace, který určuje sklon křivky grafoelementu (viz kapitola 1.2.1). Jedná se o charakter signálu, který jmenované shluky odlišuje od shluků fyziologické aktivity mozku, jak lze vidět na obrázcích 23 až 28 na straně 49.

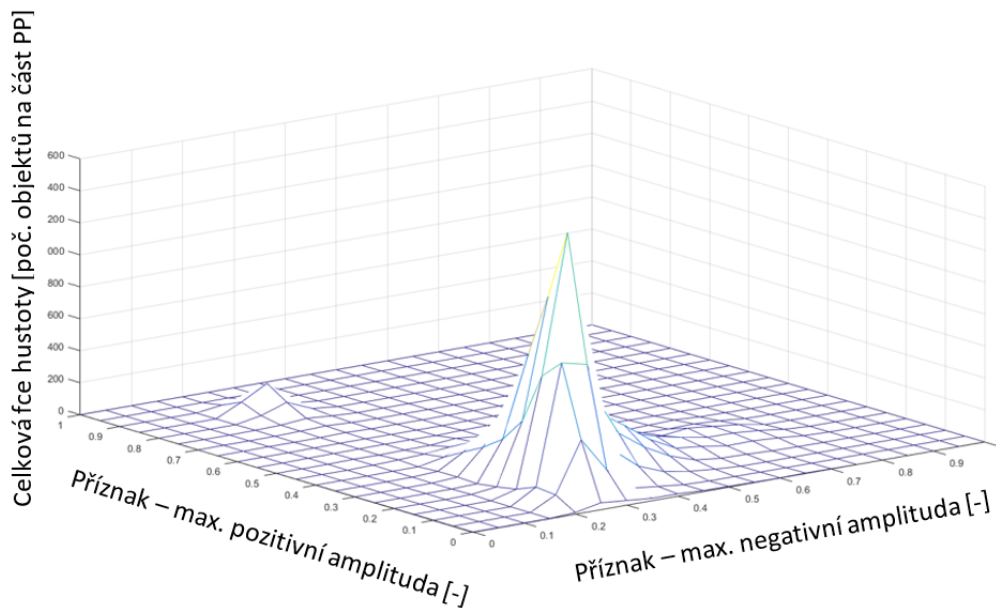
Pro provedení prvního kroku upraveného algoritmu DENCLUE jsem použil kombinaci příznaků „maximální hodnota první derivace v segmentu“ a „střední hodnota první derivace v segmentu“ (viz kapitola 1.2.1). Bylo nutné nastavit parametry, aby došlo k oddělení pouze nejvýraznějšího shluku od šumu (viz kapitola 2.4.1). Vzhledem k použití pouze 2D příznakového prostoru jsem opticky snadněji kontroloval toto nastavení. Na obrázku 33 lze vidět příklad celkové funkce hustoty pro EEG data u těchto dvou příznaků. Dále zde lze vidět oddělení šumového shluku od nejvýraznějšího shluku příznakového prostoru ve stejné kombinaci příznaků.



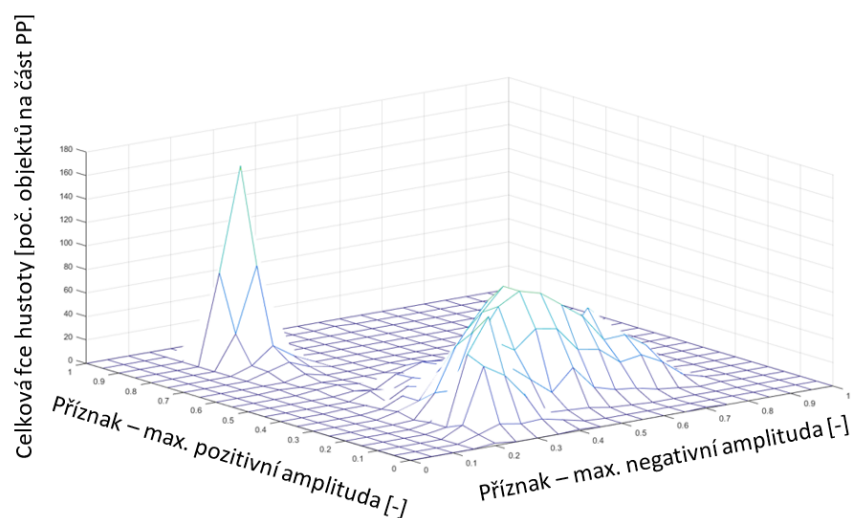
Obrázek 33: Vlevo je příklad celkové funkce hustoty pro příznaky střední hodnota první derivace v segmentu a maximální hodnota první derivace v segmentu. Celková funkce hustoty leží na ose z (má jednotku počet objektů na část příznakového prostoru), na ose x a y je zobrazen bezrozměrný příznakový prostor. Vpravo jsou na základě celkové funkce hustoty a nastavení parametrů $k = 55$ a $\xi = 40$ barevně odděleny dva shluky. Zelený shluk je šumový shluk.

Na obrázcích 34 a 35 lze vidět příklady různých kombinací 2 příznaků před a po oddělení nejvýraznějšího shluku příznakového prostoru od ostatních shluků. Z obrázků 34 a 35 je patrné, že se po oddělení nejvýraznějšího shluku zvýraznily ostatní shluky. Tyto shluky bylo možné snadněji od sebe odlišit. Toto platilo pro všechny kombinace 2 příznaků (ze sady příznaků, které se v této práci používají - viz kapitola 1.2.1). Dalo se předpokládat, že podobný efekt bude mít tato činnost i v příznakovém prostoru o 23 dimenzích složeném z těchto příznaků.

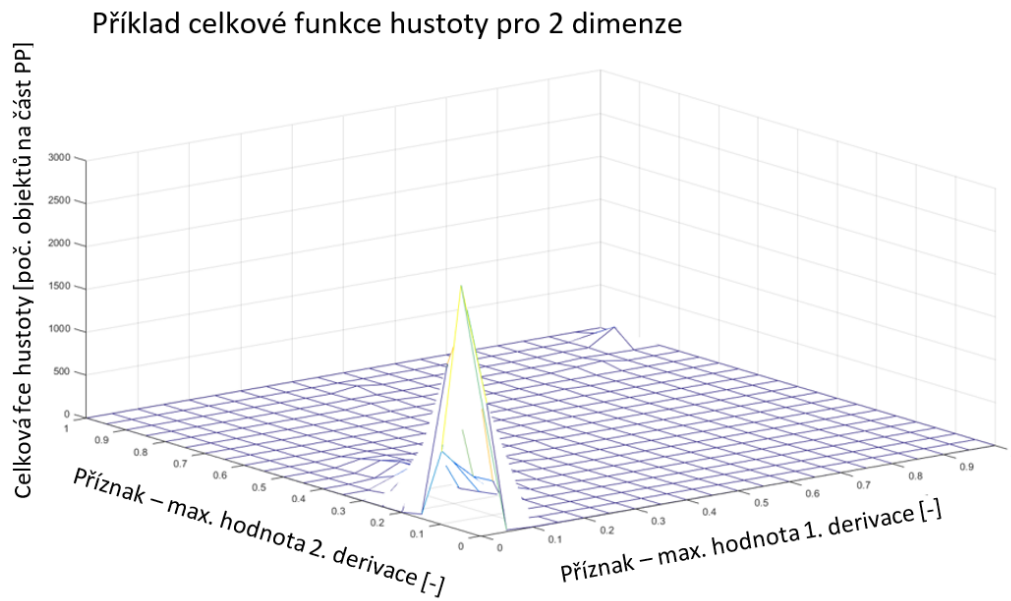
Příklad celkové funkce hustoty pro 2 dimenze



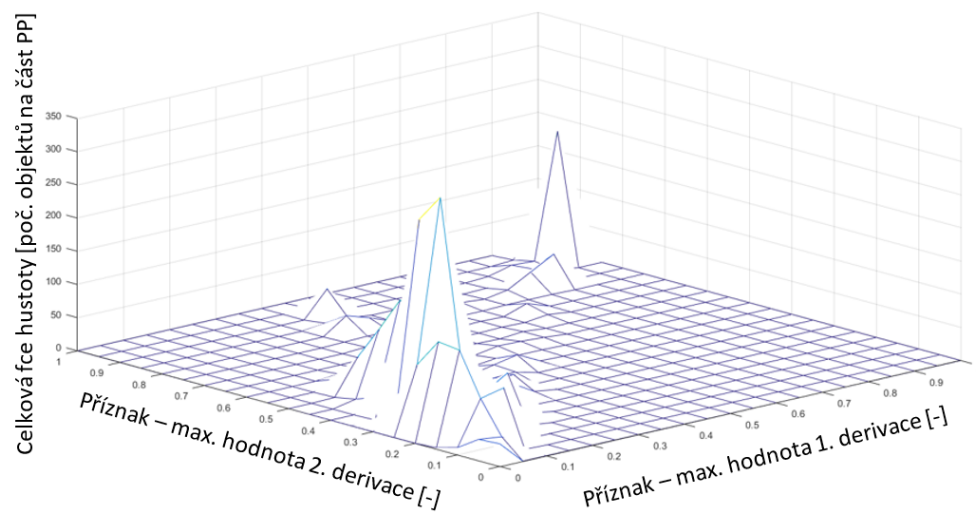
Příklad celkové funkce hustoty pro 2 dimenze po odstranění nejvýraznějšího shluku



Obrázek 34: Nahoře je příklad celkové funkce hustoty před odebráním nejvýraznějšího shluku. Dole je celkové funkce hustoty pro stejnou kombinace dvou příznaků po odebrání nejvýraznějšího shluku. Na ose z leží celková funkce hustoty (mající jednotku počet objektů na část příznakového prostoru), na ose x a y se nachází bezrozměrný příznakový prostor.



Příklad celkové fce hustoty pro 2 dimenze po odstranění nejvýraznějšího shluku



Obrázek 35: Nahoře je příklad celkové funkce hustoty před odebráním nejvýraznějšího shluku. Dole je celkové funkce hustoty pro stejnou kombinace dvou příznaků po odebrání nejvýraznějšího shluku. Na ose z leží celková funkce hustoty (mající jednotku počet objektů na část příznakového prostoru), na ose x a y se nachází bezrozměrný příznakový prostor.

Pro oddělení shluků v první fázi klasifikovaných jako šum, již nebylo potřeba používat všech 23 příznaků využívaných pro neupravený algoritmus DENCLUE. Pro tyto shluky nebyla potřebná například hodnota amplitudy signálu v beta či alfa frekvenčním pásmu.

Zmenšením počtu příznaků se výrazně zvýšila časová náročnost algoritmu (viz kapitoly 3.2.1 a 4.2).

Pro druhý krok upraveného algoritmu DENCLUE byly vybrány 3 příznaky. Prvním a druhým příznakem byly „maximální negativní amplituda v daném segmentu“ a „maximální pozitivní amplituda v daném segmentu“. Třetím příznakem byl „počet inflexních bodů“. Pro tyto vybrané kombinace příznaků byly pozorovány nejlepší výsledky u trénovacího EEG záznamu.

2.4.1 Nastavení parametrů

Vstupními parametry algoritmu DENCLUE jsou koeficient šumu ξ a koeficient hladkosti h . Ty je potřeba nastavit před spuštěním algoritmu. V našem případě byl koeficient hladkosti nahrazený parametrem počtu buněk histogramu k (viz kapitola 1.3).

Čím větší byl koeficient šumu, tím více objektů spadalo do šumového shluku. To zmenšovalo počet vytvořených shluků, jak lze vidět na grafu z obrázku 36. Pro koeficient šumu blízký se nekonečnu zůstával pouze jeden shluk obsahující všechny objekty. Jednalo se o šumový shluk. Pokud se koeficient šumu blížil hodnotě 1, vytvořilo se velké množství málo obsazených shluků. Výsledkem bylo, že stejné segmenty záznamu (objekty) patřící do stejného shluku se rozdělily do více shluků. Zmenšovala se tak senzitivita algoritmu (viz kapitola 2.3). Pokud se naopak zvýšila hodnota koeficientu šumu příliš, zvýšil se počet segmentů klasifikovaných jako šum. Toto zvýšení se dělo i u segmentů, které mezi šum nepatřily. Zároveň se snížil počet shluků, které byly výrazně zaplněny segmenty (objekty). V tomto případě do jednoho shluku patřilo více různých segmentů záznamu a snížila se tak PPV algoritmu (viz kapitola 2.3). Pro správné nastavení koeficientu šumu ξ bylo tedy potřeba především hlídat, jak moc byl obsazený šumový shluk v poměru k obsazenosti ostatních shluků. Na obrázcích 37 a 38 lze vidět, jakým způsobem ovlivňoval změněný parametr ξ automatickou klasifikaci přímo ve 2D příznakovém prostoru.



Obrázek 36: Graf zobrazující na ose y počet shluků v závislosti na měnícím se koeficientu šumu ξ , zobrazeném na ose x . Graf je vytvořen pro počet buněk histogramu $k = 20$. Data jsou získána z reálného EEG záznamu.

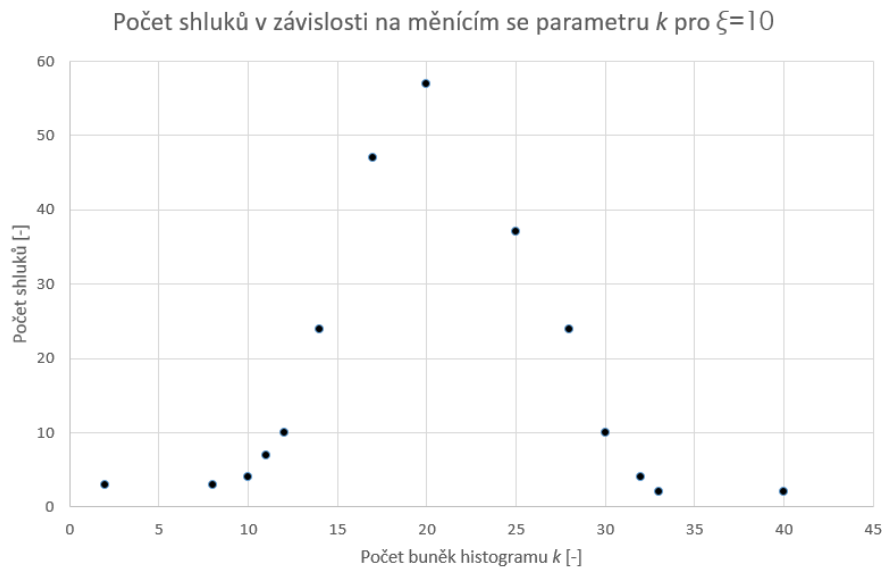


Obrázek 37: Příklad rozdělení objektů do shluků u algoritmu DENCLUE a příznaků „variabilita signálu v daném segmentu“ a „maximální pozitivní amplituda v daném segmentu“. Příklad je vytvořen pro počet buněk histogramu $k = 20$ a koeficient šumu $\xi = 10$.



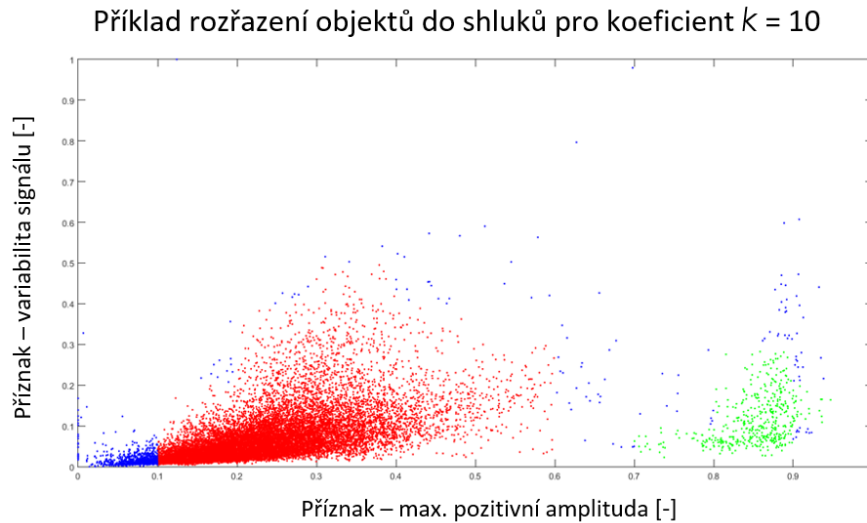
Obrázek 38: Příklad rozdělení objektů do shluků u algoritmu DENCLUE a příznaků „variabilita signálu v daném segmentu“ a „maximální pozitivní amplituda v daném segmentu“. Příklad je vytvořen pro počet buněk histogramu $k = 20$ a koeficient šumu $\xi = 100$.

Na grafu z obrázku 39 lze vidět závislost počtu shluků na počtu buněk histogramu k pro koeficient šumu $\xi = 10$. Z grafu je patrné, že s rostoucím počtem buněk histogramu roste i počet shluků až do globálního maxima. V tomto maximu převládá vliv koeficientu šumu nad parametrem počet buněk histogramu. Dále počet shluků s rostoucím parametrem k klesal. První část grafu způsobil fakt, že čím více bylo buněk v histogramu, tím větší byla různorodost zaplněnosti těchto buněk. Zároveň bylo i více lokálních maxim v celkové funkci hustoty (viz kapitola 1.3). Pokud rostl počet buněk histogramu, klesala zároveň maximální zaplněnost jednotlivých buněk. To způsobilo charakter druhé části křivky. Čím méně byly zaplněné jednotlivé buňky, tím více zde rostl počet objektů zařazených do šumového shluku. Toto se dělo na úkor ostatních shluků, které postupně zanikaly.

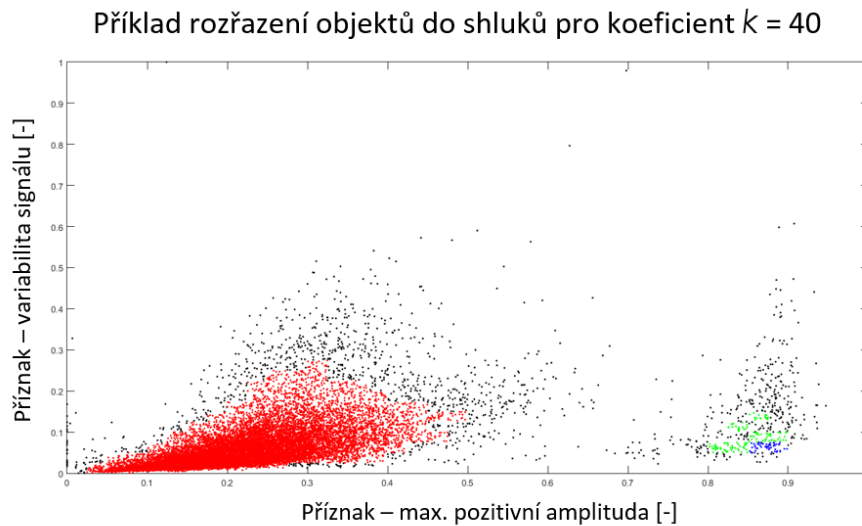


Obrázek 39: Graf zobrazující na ose y počet shluků v závislosti na měnícím se počtu buněk histogramu k , zobrazeném na ose x . Graf je vytvořen pro počet buněk histogramu $\xi = 10$. Data jsou získána z reálného EEG záznamu.

Pokud se nastavil příliš nízký parametr k , vznikl pouze malý počet buněk. V jedné buňce byly různé segmenty záznamu (objekty), což snížilo PPV algoritmu (viz kapitola 2.3). Pokud se naopak příliš zvýšil počet buněk histogramu, vzniklo více shluků, které obsahovaly stejné segmenty záznamu. To snížilo doplněk specifity algoritmu (viz kapitola 2.3). Při výrazném zvýšení počtu buněk histogramu se zvětšil počet nešumových segmentů klasifikovaných jako šum. Princip chování byl poté podobný jako u nastavení příliš velkého koeficientu šumu. Pro správné nastavení počtu buněk histogramu bylo potřeba sledovat zda se segmenty patřící do stejného shluku nerozdělovaly do více shluků. Pokud se nerozdělovaly a zároveň nebyl příliš zaplněn šumový shluk, bylo možné parametr k zvětšit. Na obrázcích 40 a 41 lze vidět, jakým způsobem ovlivňoval změněný parametr k automatickou klasifikaci přímo ve 2D příznakovém prostoru.



Obrázek 40: Příklad rozdělení objektů do shluků u algoritmu DENCLUE a příznaků „variabilita signálu v daném segmentu“ a „maximální pozitivní amplituda v daném segmentu“. Příklad je vytvořen pro koeficient šumu $\xi = 15$ a počet buněk histogramu $k = 10$.



Obrázek 41: Příklad rozdělení objektů do shluků u algoritmu DENCLUE a příznaků „variabilita signálu v daném segmentu“ a „maximální pozitivní amplituda v daném segmentu“. Příklad je vytvořen pro koeficient šumu $\xi = 15$ a počet buněk histogramu $k = 40$.

Pro automatickou klasifikaci EEG záznamu neupraveným algoritmem DENCLUE jsem určil ideální nastavení parametrů $k = 9$ a $\xi = 5$. U algoritmu DENCLUE uzpůsobeného na EEG záznam jsem pro oddělení nejvýraznější shluk vybral parametry $k = 55$ a $\xi = 40$. Vysoká hodnota koeficientu šumu a počtu buněk histogramu způsobila oddělení šumu

od jednoho výrazného shluku. Pro druhou část jsem určil za nejvhodnější nastavení parametrů $k = 11$ a $\xi = 2$. Všechny parametry jsou nastavené pro všechny EEG záznamy stejně. Díky tomu algoritmus funguje automaticky. Pokud by se nastavovaly parametry pro každý konkrétní EEG záznam zvlášť, byly by pravděpodobně získány lepší výsledky. Ovšem nastavování parametrů by vyžadovalo několik různých spuštění automatické klasifikace a bylo by tak v klinické praxi nevýhodné.

3 Výsledky

3.1 Simulovaná data

Pro testování algoritmu DENCLUE a algoritmu k-means jsem používal 4 druhy simulovaných dat. Každý druh simuloval možné rozložení objektů EEG záznamu v příznakovém prostoru. Jednalo se o data ve 2D příznakovém prostoru (viz kapitola 2.2.1). Data testovala schopnost algoritmů správně zařadit objekty pro různé kombinace shluků. Testování proběhlo na výrazně od sebe oddělených shlucích, nevýrazně oddělených shlucích, prolnutých shlucích a shlucích s různou vzájemnou vzdáleností a hustotou.

3.1.1 Kvantitativní analýza

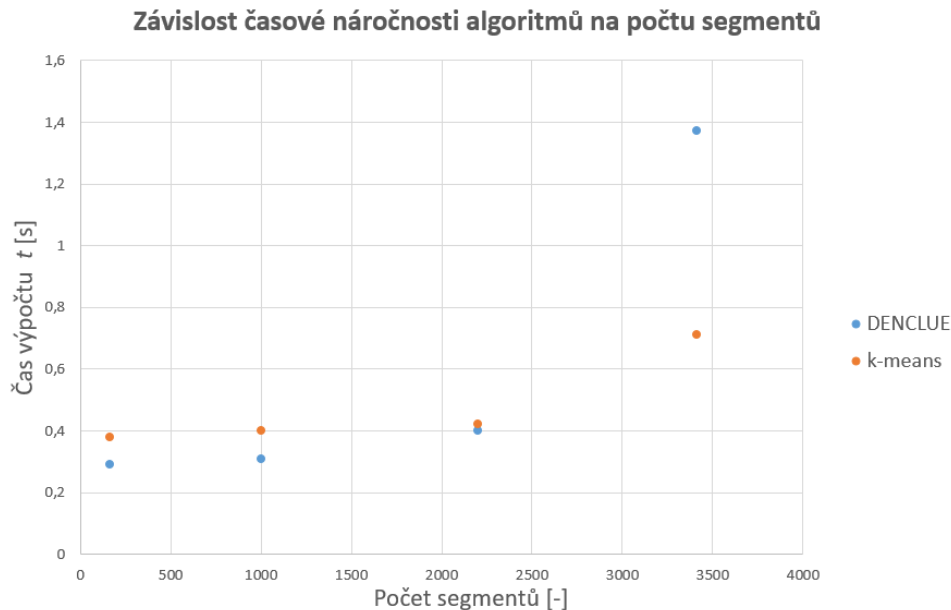
Při kvantitativní analýze jsem určoval časovou náročnost algoritmů. Časová náročnost ukazuje dobu, za kterou algoritmus provede svojí činnost a počítal jsem ji v programu MATLAB R2015a [26]. Časovou náročnost jsem počítal pro algoritmus DENCLUE neupravený na EEG záznam a algoritmus k-means. Časovou náročnost jsem získával pro různý počet segmentů simulovaných dat. Pro každou kombinaci algoritmů a počtu segmentů jsem počítal časovou náročnost pětkrát. Časovou náročnost z tabulky 4 jsem určil mediánem všech měřených časů.

V tabulce 4 lze vidět medián časové náročnosti algoritmů DENCLUE a k-means pro simulovaná data. Časová náročnost je ukázána v závislosti na počtu segmentů simulovaných dat.

Tabulka 4: Časová náročnost v sekundách u algoritmů DENCLUE a k-means pro počet segmentů simulovaných dat.

	k-means	DENCLUE
počet segmentů [-]	čas. nároč. t [s]	čas. nároč. t [s]
164	0,38	0,29
1000	0,40	0,31
2200	0,42	0,40
3412	0,71	1,37

Na obrázku 42 je vykreslen graf hodnot z tabulky 4. Graf je vykreslen pro algoritmus k-means i algoritmus DENCLUE upravený na EEG záznam.



Obrázek 42: Graf závislosti časové náročnosti algoritmů k-means (červené body) a DENCLUE (modré body) na počtu segmentů simulovaných dat.

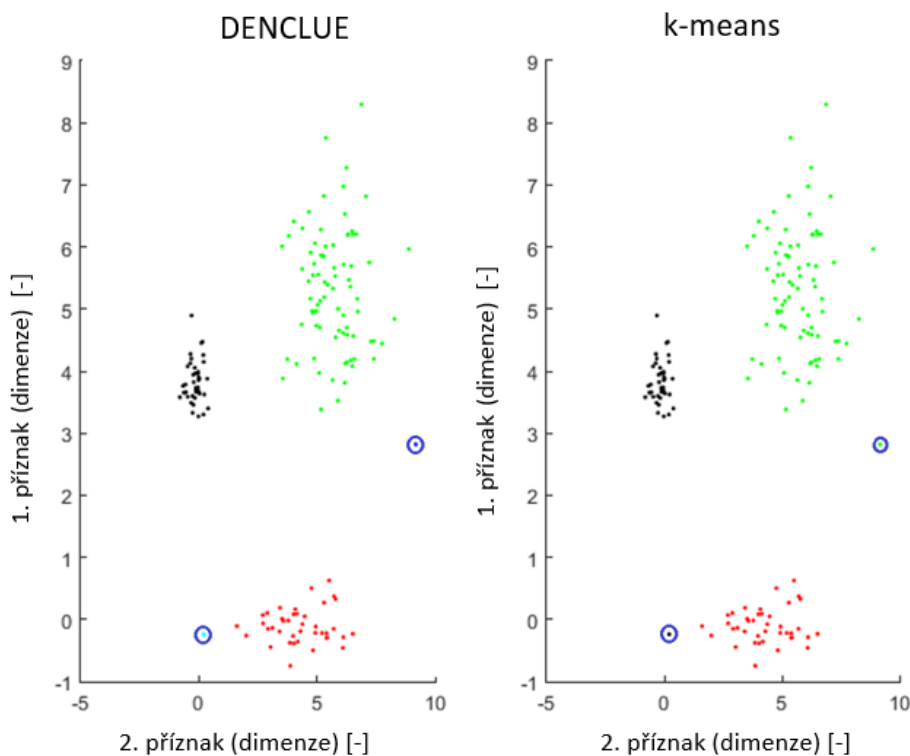
3.1.2 Kvalitativní analýza

Během kvalitativní analýzy simulovaných dat jsem porovnával správné zařazení objektů do shluků. Toto porovnání jsem prováděl pro algoritmy DENCLUE neupravený na EEG záznam a k-means. Porovnání jsem provedl prostřednictvím ROC analýzy. Konkrétně jsem pro každý typ simulovaných dat zjišťoval senzitivitu, specificitu a PPV. Jednotlivé parametry jsem počítal pro každý shluk simulovaných dat i pro simulovaná data jako celek. U algoritmu DENCLUE jsem nastavoval parametry pro konkrétní typ simulovaných dat. U algoritmu k-means jsem nastavoval počet shluků stejný se skutečným počtem shluků v simulovaných datech.

Výrazně oddělené shluky

Prvním typem simulovaných dat byly výrazně oddělené shluky. Tento typ simulovaných dat určoval, zda jsou testované algoritmy schopné rozlišit základní shluky. Na obrázku 43 lze vidět, jakým způsobem jednotlivé algoritmy rozdělily objekty do shluků.

Rozřazení objektů do shluků pro simulovaná data s oddělenými shluky



Obrázek 43: Rozřazení simulovaných dat do barevně oddělených shluků v bezrozměrném příznakovém prostoru. Vlevo je výsledek rozřazení pro algoritmus DENCLUE, vpravo je výsledek rozřazení pro algoritmus k-means. Simulovaná data byla vytvořena s výrazně oddělenými shluky. Modře jsou zakroužkované objekty, které zařazovaly dané algoritmy špatně.

Výsledky rozdělení objektů do shluků z obrázku 43 lze vidět v tabulce 5. V tabulce jsou ukázány parametry senzitivita, specificita a PPV pro jednotlivé shluky i pro celá simulovaná data.

Tabulka 5: Specificita (SP), senzitivita (SZ) a prediktivní hodnota pozitivního testu (PPV) pro algoritmy k-means a DENCLUE. Hodnoty jsou ukázány pro jednotlivé shluky simulovaných dat i pro celá simulovaná data s výrazně oddělenými shluky.

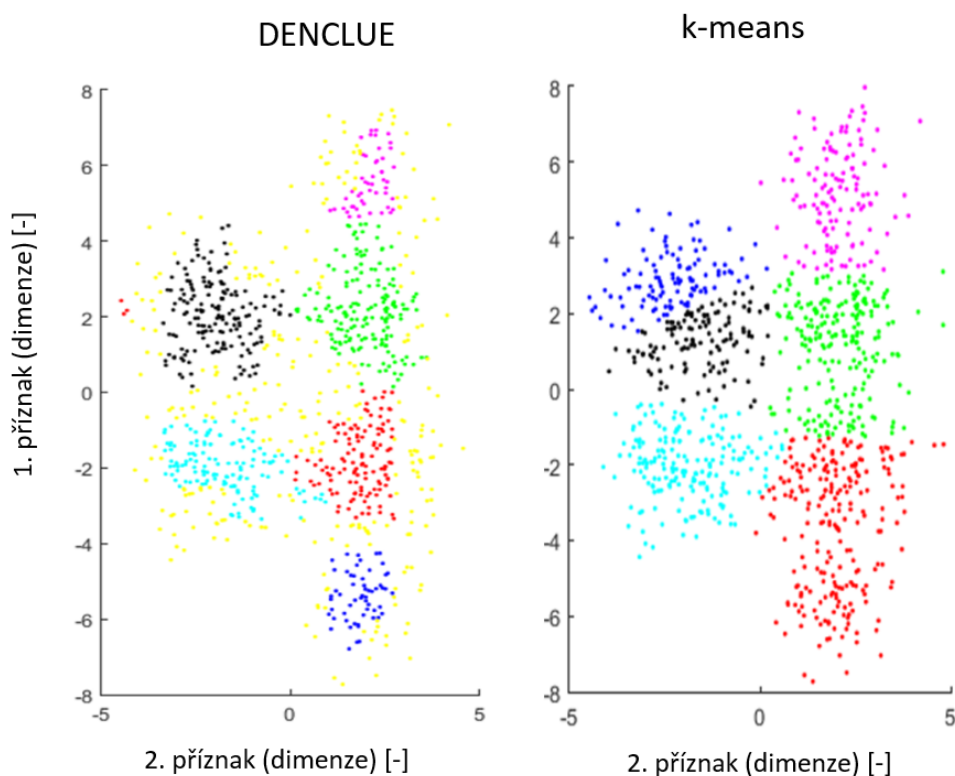
Oddělené shluky	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Shluk 1	1,00	1,00	1,00	1,00	1,00	1,00
Shluk 2	1,00	0,98	1,00	1,00	0,98	1,00
Shluk 3	1,00	1,00	1,00	1,00	0,99	1,00
celkem	1,00	0,99	1,00	1,00	0,99	1,00

Z tabulky je patrné, že oba algoritmy měly stejnou specificitu, senzitivitu a PPV pro celá simulovaná data. Pro senzitivitu a PPV zde byly nejvyšší možné hodnoty. Specificita se blížila číslu 1. Nepatrný rozdíl mezi oběma algoritmy se nacházel pouze u senzitivity druhého shluku.

Nevýrazně oddělené shluky

Druhý typ simulovaných dat obsahoval šest shluků, které byly umístěny blízko sobě. Tato data testovala schopnost algoritmů oddělit i částečně prolnuté shluky. Na obrázku 44 lze pozorovat rozdělení objektů těchto dat do shluků u jednotlivých algoritmů.

Rozřazení objektů do shluků pro simulovaná data s neoddělenými shluky



Obrázek 44: Rozřazení simulovaných dat do barevně oddělených shluků v bezrozměrném příznakovém prostoru. Vlevo je výsledek rozřazení pro algoritmus DENCLUE, vpravo je výsledek rozřazení pro algoritmus k-means. Simulovaná data byla vytvořena s nevýrazně oddělenými shluky.

Statistické zpracování rozdělení objektů z obrázku 44 lze vidět v tabulce 6. V této tabulce je zapsaná senzitivita, specificita a PPV pro algoritmy DENCLUE a k-means. Tyto hodnoty byly vytvořeny pro jednotlivé shluky i pro celá simulovaná data.

Tabulka 6: Specificita (SP), senzitivita (SZ) a prediktivní hodnota pozitivního testu (PPV) pro algoritmy k-means a DENCLUE. Hodnoty jsou ukázány pro jednotlivé shluky simulovaných dat i pro celá simulovaná data s nevýrazně oddělenými shluky.

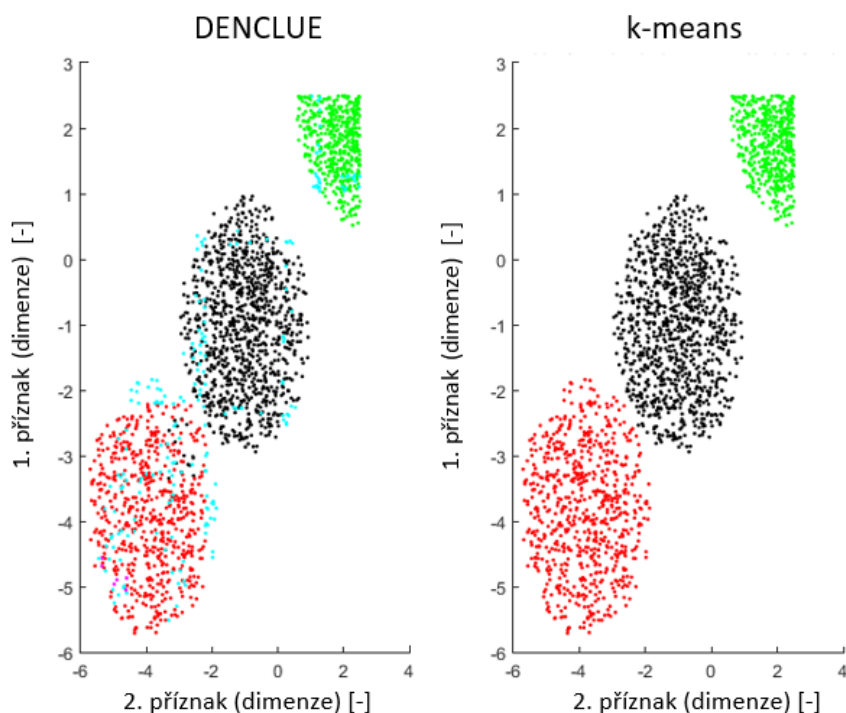
Neoddělené shluky	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV [-]
Shluk 1	0,99	0,52	0,94	1,00	0,74	1,00
Shluk 2	0,99	0,85	0,95	1,00	0,95	0,98
Shluk 3	0,99	0,91	0,95	0,99	0,65	0,93
Shluk 4	0,91	0,53	0,66	1,00	0,63	1,00
Shluk 5	alg. nenalezl shluk			1,00	0,59	1,00
Shluk 6	1,00	0,82	1,00	1,00	0,61	1,00
celkem	0,96	0,70	0,79	1,00	0,71	0,98

Z tabulky 6 lze vyčíst, že hodnoty senzitivity pro jednotlivé shluky se u algoritmu k-means pohybovaly od 0,52 do 0,91. U algoritmu DENCLUE se hodnoty senzitivity pro jednotlivé shluky pohybovaly od 0,59 do 0,95. Nejnižší specificita se vyskytovala u shluku 4 pro algoritmus k-means. Její hodnota byla 0,91. Shluk 5 algoritmus k-means vůbec nenašel, namísto toho rozdělil shluk 1. Celkově měly oba algoritmy nižší senzitivitu (algoritmus k-means 0,70, algoritmus DENCLUE 0,71). Výraznější rozdíl byl pro celkovou PPV obou algoritmů. K-means měl tuto hodnotu 0,7 a DENCLUE 0,98.

Shluky s různou vzájemnou vzdáleností

Třetí typ simulovaných dat obsahoval tři shluky. Dva shluky byly umístěny blízko sebe. Třetí shluk byl umístěn dále od těchto dvou a měl vyšší hustotu. Na obrázku 45 lze vidět rozdělení objektů těchto dat do shluků. Rozdělení do shluků je ukázáno pro algoritmy DENCLUE a k-means.

Rozřazení objektů do shluků pro simulovaná data s různou vzdáleností shluků



Obrázek 45: Rozřazení simulovaných dat do barevně oddělených shluků v bezrozměrném příznakovém prostoru. Vlevo je výsledek rozřazení pro algoritmus DENCLUE, vpravo je výsledek rozřazení pro algoritmus k-means. Simulovaná data byla vytvořena se shluky o různé vzájemné vzdálenosti a hustotě.

Statistické zpracování rozdělení objektů do shluků lze vidět v tabulce 7. V této tabulce je vidět parametry senzitivitu, specificitu a PPV pro jednotlivé shluky i pro celá simulovaná data.

Tabulka 7: Specificita (SP), senzitivita (SZ) a prediktivní hodnota pozitivního testu (PPV) pro algoritmy k-means a DENCLUE. Hodnoty jsou ukázány pro jednotlivé shluky simulovaných dat i pro celá simulovaná data se shluky o různé vzájemné vzdálenosti a hustotě.

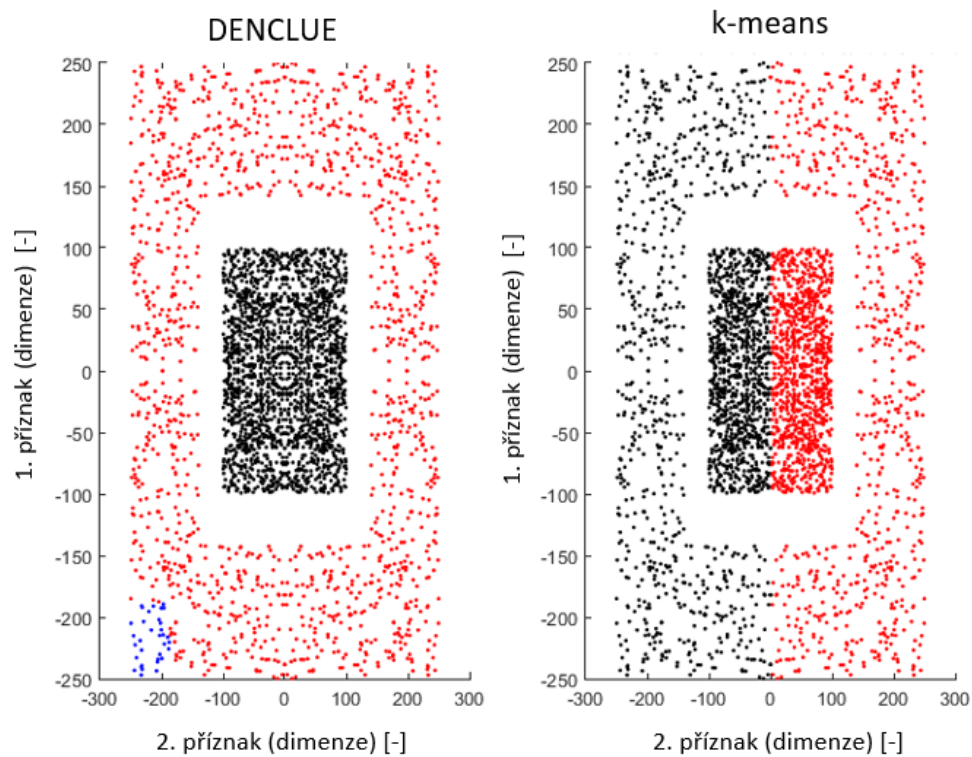
Různá vzdál. shluků	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV [-]
Shluk 1	1,00	1,00	1,00	1,00	0,94	1,00
Shluk 2	1,00	1,00	1,00	0,99	0,94	0,99
Shluk 3	1,00	1,00	1,00	1,00	0,82	1,00
celkem	1,00	1,00	1,00	1,00	0,90	0,99

Z tabulky 7 je patrné, že algoritmus k-means měl senzitivitu, specificitu a PPV rovnou jedné pro všechny shluky i celková simulovaná data. Algoritmus DENCLUE měl senzitivitu pro celá simulovaná data 0,90.

Prolnuté shluky

Čtvrtý typ simulovaných dat reprezentoval dva prolnuté shluky. Na těchto simulovaných datech se testovaly lepší vlastnosti hustotního algoritmu DENCLUE oproti algoritmu na bázi vzdálenosti k-means. Na obrázku 46 lze vidět rozdělení objektů z těchto dat do shluků pro algoritmus DENCLUE a pro algoritmus k-means.

Rozřazení objektů do shluků pro simulovaná data s prolnutými shluky



Obrázek 46: Rozřazení simulovaných dat do barevně oddělených shluků v bezrozměrném příznakovém prostoru. Vlevo je výsledek rozřazení pro algoritmus DENCLUE, vpravo je výsledek rozřazení pro algoritmus k-means. Simulovaná data byla vytvořena s prolnutými shluky.

Statistické zpracování výsledků rozdělení objektů do shluků lze vidět v tabulce 8. V tabulce jsou uvedeny parametry senzitivita, specifita a PPV pro všechny shluky a celá simulovaná data.

Tabulka 8: Specifita (SP), senzitivita (SZ) a prediktivní hodnota pozitivního testu (PPV) pro algoritmy k-means a DENCLUE. Hodnoty jsou ukázány pro jednotlivé shluky simulovaných dat i pro celá simulovaná data s prolnutými shluky.

Prolnuté shluky	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Shluk 1	0,58	0,59	0,59	1,00	1,00	1,00
Shluk 2	0,58	0,58	0,58	1,00	0,98	1,00
celkem	0,58	0,59	0,59	1,00	1,00	1,00

Z tabulky 8 je patrné, že algoritmus DENCLUE měl pro celá simulovaná data senzitivitu, specifitu i PPV rovnou jedné. Algoritmus k-means měl oproti algoritmu DENCLUE hodnoty senzitivity, specifity a PPV výrazně nižší. Hodnota specifity byla pro celá simulovaná data u algoritmu k-means 0,58.

3.2 Reálné EEG záznamy

Pro porovnání testovaných algoritmů jsem využíval reálné EEG záznamy. EEG záznamy byly naměřeny v Nemocnici Na Bulovce na 15 subjektech. Na reálných EEG záznamech jsem prováděl kvantitativní a kvalitativní analýzu porovnávaných algoritmů. Při kvantitativní analýze jsem určoval časovou náročnost porovnávaných algoritmů. Při kvalitativní analýze jsem statisticky vyhodnocoval správnost zařazení segmentů EEG záznamu do shluků.

Správnost zařazení segmentů EEG záznamu do shluků určovali dva experti. Pro velký počet segmentů EEG záznamů nebylo možné, aby experti analyzovali celých 15 EEG záznamů. Proto jsem kvalitativní analýzu rozdělil na analýzu celého EEG záznamu a analýzu více záznamů. Analýzu celého EEG záznamu jsem provedl na třech kompletních EEG záznamech. Analýzu více EEG záznamů jsem provedl na 12-ti EEG záznamech, kde jsem náhodně vybral 50 segmentů v každém záznamu.

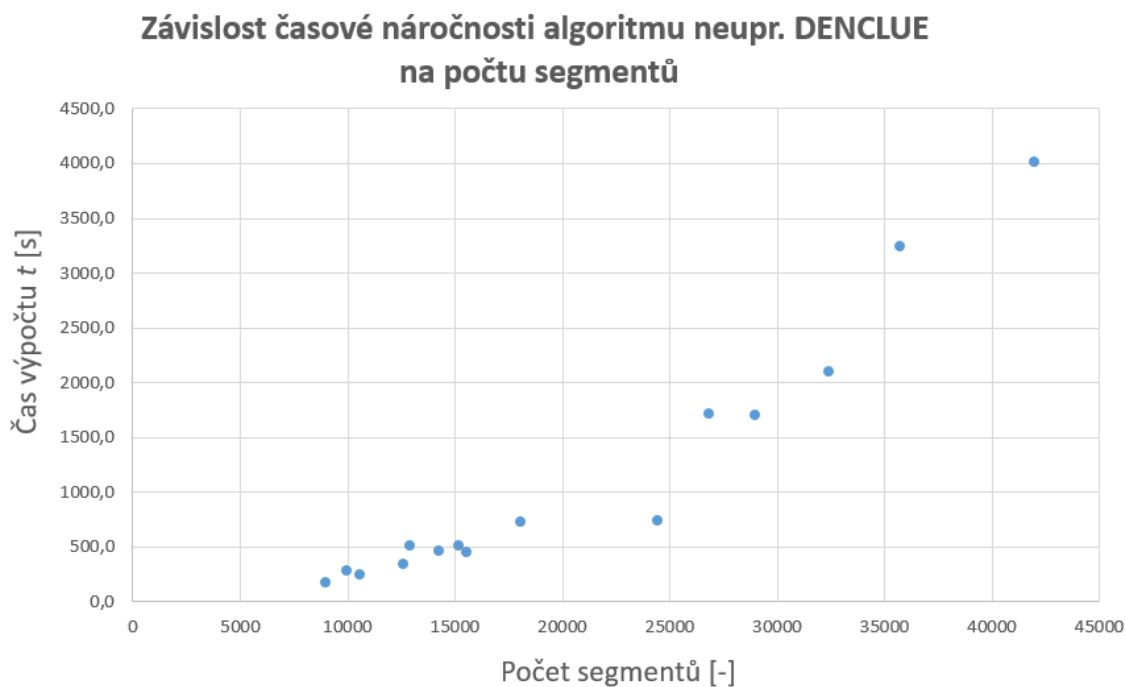
3.2.1 Kvantitativní analýza

Během kvantitativní analýzy reálných EEG záznamů jsem porovnával časovou náročnost algoritmů k-means, DENCLUE neupraveného na EEG záznam a DENCLUE upraveného na EEG záznam. Časová náročnost ukazuje čas, za který vykonají algoritmy svojí činnost. Ten jsem zjišťoval pro 15 EEG záznamů o různém počtu segmentů. Výsledná časová náročnost byla určena jako medián z 5 časových náročností daného algoritmu pro daný počet segmentů EEG záznamu.

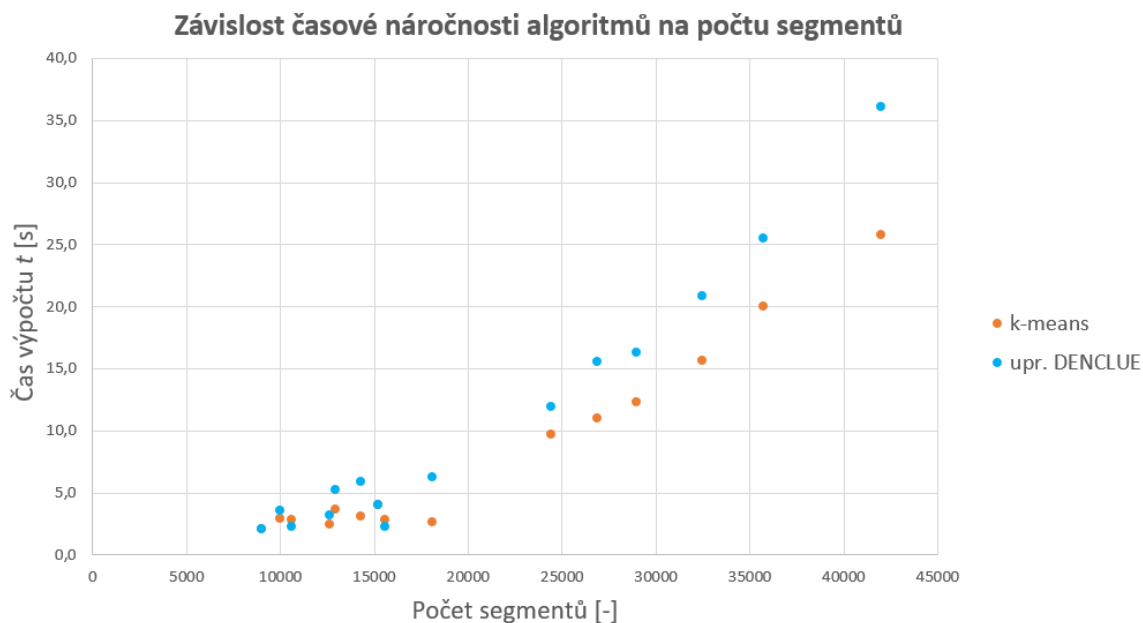
V tabulce 9 lze pozorovat výsledky časové náročnosti jednotlivých algoritmů pro různý počet segmentů EEG záznamů. Na obrázku 47 lze vidět graf závislosti časové náročnosti algoritmu DENCLUE neupraveného na EEG záznam na počtu segmentů EEG záznamu. Na obrázku 48 lze vidět graf závislosti časové náročnosti algoritmů k-means a DENCLUE upraveného na EEG záznam na počtu segmentů EEG záznamu.

Tabulka 9: Časová náročnost u algoritmů DENCLUE upraveného na EEG záznam, DENCLUE neupraveného na EEG záznam a k-means pro počet segmentů reálných EEG záznamů. Časová náročnost je uvedena v sekundách.

	k-means	neup. DENCLUE	upr. DENCLUE
počet segmentů [-]	čas. nároč. t [s]	čas. nároč. t [s]	čas. nároč. t [s]
8957	2,1	173,9	2,1
9944	3,0	280,4	3,6
10554	2,9	250,6	2,3
12594	2,5	335,6	3,2
12884	3,7	506,3	5,3
14270	3,1	464,4	5,9
15176	4,1	511,7	4,1
15557	2,9	450,1	2,3
18049	2,7	721,3	6,3
24419	9,7	733,8	12,0
26840	11,0	1711,1	15,6
28963	12,3	1701,3	16,3
32431	15,7	2103,7	20,9
35712	20,1	3241,3	25,5
41963	25,8	4012,7	36,1



Obrázek 47: Graf závislosti časové náročnosti algoritmu DENCLUE neupraveného na EEG záznam na počtu segmentů reálných EEG záznamů.



Obrázek 48: Graf závislosti časové náročnosti algoritmů k-means (červené body) a DENCLUE upraveného na EEG záznam (modré body) na počtu segmentů reálných EEG záznamů.

3.2.2 Kvalitativní analýza celého EEG záznamu

Při kvalitativní analýze celého EEG záznamu jsem porovnával algoritmy k-means, DENCLUE neupravený na EEG záznam a DENCLUE upravený na EEG záznam. Všechny tři algoritmy jsem porovnával na třech kompletních EEG záznamech (od subjektů 1 až 3). Pro statistické vyhodnocení správnosti zařazení EEG segmentů do shluků jsem používal ROC analýzu. Konkrétně jsem využíval specifitu, senzitivitu a PPV. Pomocí této kvalitativní analýzy jsem porovnal účinnost testovaných algoritmů na automatickou klasifikaci celého EEG záznamu.

Subjekt 1

V tabulce 10 lze vidět počet segmentů patřících do hledaných shluků EEG záznamu a jejich procentuální zastoupení v EEG záznamu. Počet je ukázán pro správné zařazení segmentů EEG záznamu do shluků u subjektu 1. V tabulce 11 lze vidět senzitivitu, specifitu a PPV pro EEG záznam stejného subjektu. Tyto parametry jsou v tabulce ukázány u hledaných shluků pro algoritmy k-means, DENCLUE upravený na EEG záznam a DENCLUE neupravený na EEG záznam.

Tabulka 10: Počet segmentů v každém hledaném shluku a celkové EEG záznamu subjektu 1. Procentuální zastoupení segmentů těchto shluků v celém EEG záznamu. Nejsou započítávány segmenty vzniklé špatnou adaptivní segmentací.

Subjekt 1		
Hledané shluky	Počet segmentů [-]	Proc. zastoupení [%]
Fyz. akt.	30280	95,1
Epilep.	206	0,6
Pomal oč.	292	0,9
EMG	1020	3,2
Impulzní	3	0,0
Rov. čáry	49	0,2
celkem	31850	-

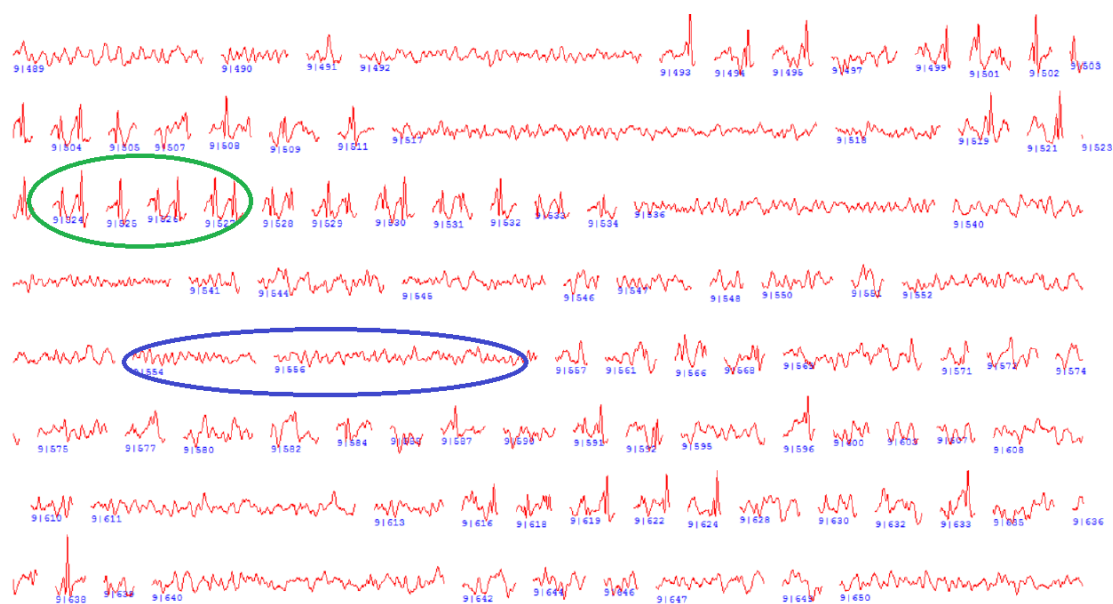
Tabulka 11: Specificita (SP), senzitivita (SZ) a prediktivní hodnota pozitivního testu (PPV) pro algoritmy k-means, DENCLUE neupravený na EEG záznam a DENCLUE upravený na EEG záznam. Vše je získáno pro shluky fyziologické aktivity mozku, epileptických grafoelementů, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu, impulzních artefaktů a rovných čar u subjektu 1.

Subjekt 1	k-means			neupr. DENCLUE			upr. DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	1,00	0,32	1,00	0,98	0,74	0,99	0,96	1,00	0,98
Epilep.	0,90	0,90	0,05	alg. nenalezl shluk			1,00	0,11	0,96
Pomal. oč.	0,87	0,47	0,03	alg. nenalezl shluk			1,00	0,05	0,21
EMG	1,00	0,86	1,00	0,80	0,92	0,80	1,00	0,91	0,84
Impulzní	alg. nenalezl shluk			alg. nenalezl shluk			alg. nenalezl shluk		
Rov. čáry	0,93	0,37	0,01	1,00	0,29	1,00	1,00	0,96	0,68

Z tabulky 10 lze vyčíst, že EEG záznam obsahoval 95,1 % segmentů fyziologické aktivity mozku. Nejvýraznější zastoupení poté měly segmenty EMG artefaktů a zašuměného signálu. Byly zde pouze 3 segmenty řazeny do shluku impulzních artefaktů. EEG záznam obsahoval 0,6 % segmentů epileptických grafoelementů.

Z tabulky 11 lze vyčíst, že ani jeden algoritmus nenalezl shluk impulzních artefaktů. Algoritmus DENCLUE neupravený na EEG záznam navíc nenalezl ani shluky epileptických grafoelementů a pomalých očních artefaktů. Dále je z tabulky patrné, že shluk epileptických grafoelementů u algoritmu k-means měl PPV 0,05 a senzitivitu 0,90. Shluk epileptických grafoelementů u algoritmu DENCLUE upraveného na EEG záznam měl senzitivitu 0,11. Algoritmus DENCLUE upravený na EEG záznam měl PPV u shluku EMG artefaktů a zašuměného signálu 0,84. Stejný algoritmus měl specificitu u shluku fyziologické aktivity mozku 0,96. Senzitivitu u shluku pomalých očních artefaktů měl tentýž algoritmus 0,05.

Na obrázcích 49 a 50 lze vidět příklady vzniklých shluků u subjektu 1 pro algoritmy k-means a DENCLUE upravený na EEG záznam. Obrázek 49 ukazuje příklad shluku epileptických grafoelementů vytvořeného algoritmem k-means. Na obrázku 50 lze vidět příklad shluku EMG artefaktů a zašuměného signálu vytvořeného algoritmem DENCLUE upraveným na EEG záznam.



Obrázek 49: Příklad shluku epileptických grafoelementů vytvořeného algoritmem k-means. Příklad je vzat z EEG záznamu subjektu 1. Zeleně je zde zakroužkován příklad epileptických grafoelementů a modře je zakroužkován příklad fyziologické aktivity mozku, která do tohoto shluku nepatří. Zobrazeno v programu WF [5].



Obrázek 50: Příklad shluku EMG artefaktů a zašuměného signálu vytvořeného algoritmem DENCLUE upraveným na EEG záznam. Příklad je vzat z EEG záznamu subjektu 1. Zeleně je zde zakroužkován příklad EMG artefaktů a modře je zakroužkován příklad fyziologické aktivity mozku, která do shluku nepatří. Zobrazeno v programu WF [5].

Subjekt 2

V tabulce 12 lze vidět počet segmentů patřících do hledaných shluků EEG záznamu a jejich procentuální zastoupení v celém EEG záznamu. Počet je ukázán pro správné zařazení segmentů do shluků u EEG záznamu subjektu 2. V tabulce 13 lze vidět senzitivitu, specificitu a PPV pro EEG záznam stejného subjektu. Tyto hodnoty jsou v tabulce ukázány pro algoritmy k-means, DENCLUE upravený na EEG záznam a DENCLUE neupravený na EEG záznam a pro každý hledaný shluk.

Tabulka 12: Počet segmentů v každém hledaném shluku a celkovém EEG záznamu subjektu 2. Procentuální zastoupení segmentů v celém EEG záznamu. Nejsou započítávány segmenty vzniklé špatnou adaptivní segmentací.

Subjekt 2		
Hledané shluky	Počet segmentů [-]	Proc. zastoupení [%]
Fyz. akt.	35190	84,4
Epilep.	5005	12,0
Pomal oč.	228	0,5
EMG	733	1,8
Impulzní	506	1,2
Rov. čáry	12	0,0
celkem	41674	

Tabulka 13: Specificita (SP), senzitivita (SZ) a prediktivní hodnota pozitivního testu (PPV) pro algoritmy k-means, DENCLUE neupravený na EEG záznam a DENCLUE upravený na EEG záznam. Vše je získáno pro shluky fyziologické aktivity mozku, epileptických grafoelementů, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu, impulzních artefaktů a rovných čar u subjektu 2.

Subjekt 2	k-means			neupr. DENCLUE			upr. DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV [-]
Fyz. akt.	0,99	0,49	1,00	0,99	0,62	1,00	0,44	0,98	0,90
Epilep.	0,97	0,71	0,76	0,60	0,98	0,25	1,00	0,18	0,96
Pomal. oč.	0,72	0,76	0,01	alg. nenalezl shluk			alg. nenalezl shluk		
EMG	1,00	0,99	0,96	alg. nenalezl shluk			0,97	0,82	0,35
Impulzní	1,00	0,96	1,00	alg. nenalezl shluk			1,00	0,90	1,00
Rov. čáry	alg. nenalezl shluk			1,00	0,83	1,00	1,00	1,00	1,00

Z tabulky 12 vychází, že segmenty fyziologické aktivity mozku v EEG záznamu subjektu 2 měly zastoupení 84,4%. Druhé největší zastoupení zde měly shluky epileptických

grafoelementů, konkrétně 12,0 %. Nejméně se v EEG záznamu vyskytoval shluk rovných čar.

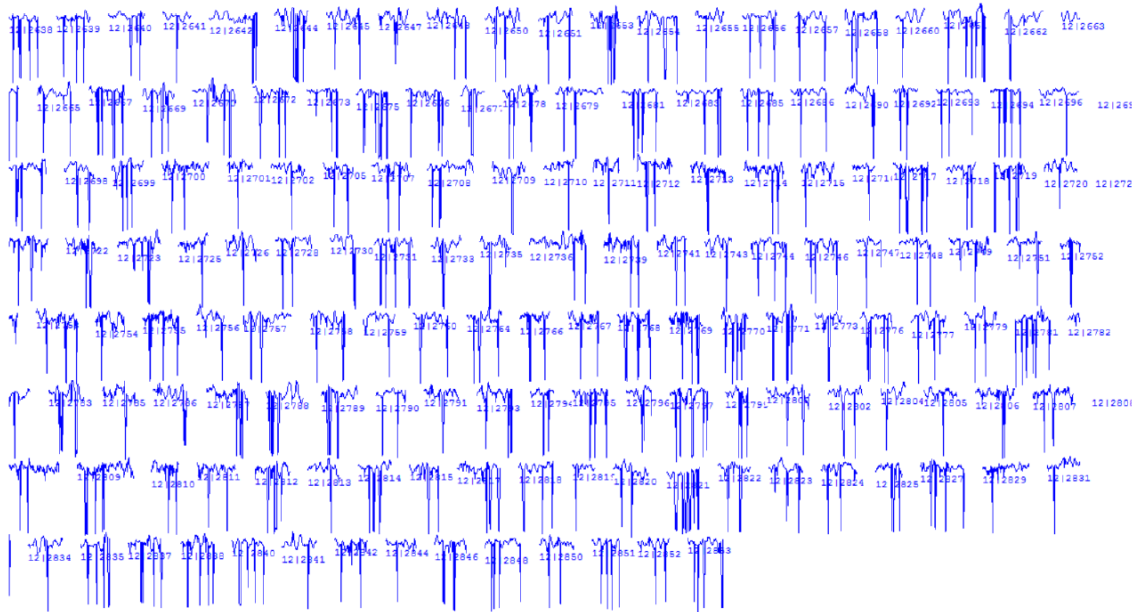
V tabulce 13 lze vidět, že ani jeden z algoritmů DENCLUE nenalezl shluk pomalých očních artefaktů. Algoritmus DENCLUE neupravený na EEG záznam navíc nenalezl ani shluky EMG artefaktů a zašuměného signálu a impulzních artefaktů. Algoritmus k-means nenalezl shluk rovných čar. Algoritmus DENCLUE upravený na EEG záznam měl senzitivitu, specificitu a PPV shluku rovných čar 1,00. Dále měl tento algoritmus senzitivitu shluku epileptických grafoelementů 0,18 a senzitivitu shluku fyziologické aktivity mozku 0,98. Oproti tomu algoritmus k-means měl senzitivitu shluku fyziologické aktivity mozku 0,49 a senzitivitu shluku epileptických grafoelementů 0,71. Algoritmus k-means a DENCLUE upravený na EEG záznam měly senzitivitu pro shluk impulzních artefaktů nad 0,89 a PPV pro stejný shluk 1,00.

Na obrázcích 51 a 52 lze vidět příklady vzniklých shluků u subjektu 2 pro algoritmy k-means a DENCLUE upravený na EEG záznam. Na obrázku 51 lze vidět příklad shluku rovných čar vytvořeného algoritmem DENCLUE upraveným na EEG záznam. Obrázek 52 ukazuje příklad shluku impulzních artefaktů vytvořeného algoritmem k-means.



1|954 2|1085 3|1065 4|1046 5|1094 7|1060 8|1047 10|972 11|991 12|1050 13|1006 15|1001

Obrázek 51: Příklad shluku rovných čar vytvořeného algoritmem DENCLUE upraveným na EEG záznam. Všechny segmenty patří do shluku. Příklad je vzat z EEG záznamu subjektu 2. Zobrazeno v programu WF [5].



Obrázek 52: Příklad shluku impulzních artefaktů vytvořeného algoritmem k-means. Všechny segmenty patří do shluku. Příklad je vzat z EEG záznamu subjektu 2. Zobrazeno v programu WF [5].

Subjekt 3

V tabulce 14 lze vidět počet segmentů patřících do hledaných shluků EEG záznamu a jejich procentuální zastoupení v celém EEG záznamu. Počet je ukázán pro správné zařazení segmentů v EEG záznamu subjektu 3. V tabulce 15 lze vidět senzitivitu, specificitu a PPV pro EEG záznam stejného subjektu. Tyto hodnoty jsou v tabulce ukázány pro algoritmy k-means, DENCLUE upravený na EEG záznam a DENCLUE neupravený na EEG záznam napříč všemi hledanými shluky.

Tabulka 14: Počet segmentů u každého hledaného shluku a celkového EEG záznamu subjektu 3. Procentuální zastoupení segmentů v celém EEG záznamu. Nejsou započítávány segmenty vzniklé špatnou adaptivní segmentací.

Subjekt 3		
Hledané shluky	Počet segmentů [-]	Proc. zastoupení [%]
Fyz. akt.	17252	95,6
Epilep.	554	3,1
Pomal. oč.	19	0,1
EMG	219	1,2
Impulzní	0	0,0
Rov. čáry	0	0,0
celkem	18044	

Tabulka 15: Specificita (SP), senzitivita (SZ) a prediktivní hodnota pozitivního testu (PPV) pro algoritmy k-means, DENCLUE neupravený na EEG záznam a DENCLUE upravený na EEG záznam. Vše je získáno pro shluky fyziologické aktivity mozku, epileptických grafoelementů, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu, impulzních artefaktů a rovných čar u subjektu 3.

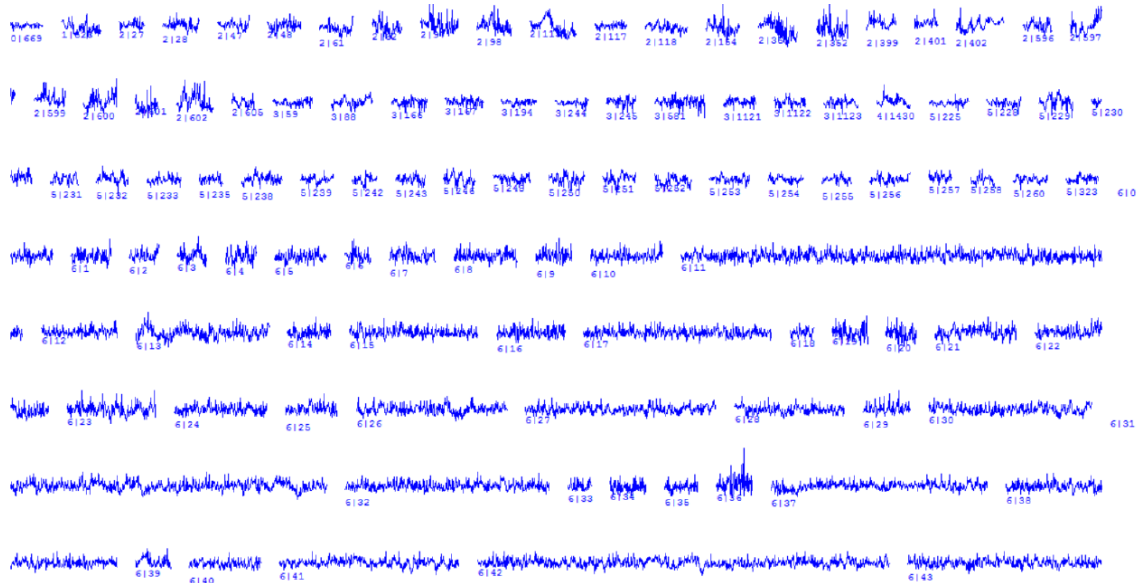
Subjekt 3	k-means			neupr. DENCLUE			upr. DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV [-]
Fyz. akt.	1,00	0,28	1,00	1,00	0,36	1,00	0,64	0,93	0,98
Epilep.	0,99	0,96	0,81	0,95	1,00	0,42	0,99	0,35	0,50
Pomal. oč.	0,94	0,70	0,01	alg. nenalezl shluk			alg. nenalezl shluk		
EMG	1,00	0,96	0,99	alg. nenalezl shluk			0,97	0,79	0,28

Z tabulky 14 lze vyčíst, že segmenty fyziologické aktivity mozku měly 95,6% zastoupení v EEG záznamu subjektu 3. Segmenty shluku epileptických grafoelementů měly v EEG záznamu zastoupení 3,1%. Segmentů shluku pomalých očních artefaktů bylo v EEG záznamu 19. Segmenty shluků impulzních artefaktů a rovných čar se v EEG záznamu nevyskytovaly vůbec.

V tabulce 15 lze vidět, že ani jeden algoritmus DENCLUE nenalezl shluk pomalých očních artefaktů. Algoritmus DENCLUE neupravený na EEG záznam nenalezl ani shluk EMG artefaktů a zašuměného signálu. Algoritmus k-means měl u shluku pomalých očních artefaktů senzitivitu 0,70 a PPV 0,01. Algoritmus DENCLUE upravený na EEG záznam

měl u shluku EMG artefaktů a zašuměného signálu senzitivitu 0,79 a PPV 0,28. U stejného shluku měl algoritmus k-means PPV 0,99.

Na obrázcích 53 a 54 lze vidět příklady vzniklých shluků u subjektu 3 pro algoritmy k-means a DENCLUE upravený na EEG záznam. Obrázek 53 ukazuje příklad shluku epileptických grafoelementů vytvořeného algoritmem k-means. Na obrázku 54 lze vidět příklad shluku EMG artefaktů a zašuměného signálu vytvořeného algoritmem DENCLUE upraveným na EEG záznam.



Obrázek 53: Příklad shluku EMG artefaktů a zašuměného signálu vytvořeného algoritmem k-means. Všechny segmenty patří do shluku. Příklad je vzat z EEG záznamu subjektu 3. Zobrazeno v programu WF [5].



Obrázek 54: Příklad shluku EMG artefaktů a zašuměného signálu vytvořeného algoritmem DENCLUE upraveným na EEG záznam. Příklad je vzat z EEG záznamu subjektu 3. Zeleně je zde zakroužkován příklad EMG artefaktů a modře je zakroužkován příklad epileptických grafoelementů, které do shluku nepatří. Zobrazeno v programu WF [5].

Subjekty 1 až 3

V tabulce 16 lze vidět počet segmentů patřících do hledaných shluků EEG záznamů a jejich procentuální zastoupení v EEG záznamech. Počet je ukázán pro správné zařazení segmentů. Tabulka je vytvořena pro EEG záznamy subjektů 1 až 3 a jsou v ní součty segmentů z EEG záznamů těchto subjektů. V tabulce 13 lze vidět senzitivitu, specifitu a PPV pro EEG záznamy stejných subjektů. Tyto hodnoty jsem získal jako medián z hodnot pro subjekty 1 až 3. Hodnoty jsou v tabulce ukázány pro algoritmy k-means, DENCLUE upravený na EEG záznam a DENCLUE neupravený na EEG záznam a pro každý hledaný shluk.

Tabulka 16: Počet segmentů u každého hledaného shluku a celkových EEG záznamů subjektů 1, 2 a 3. Procentuální zastoupení segmentů v EEG záznamech. Nejsou započítávány segmenty vzniklé špatnou adaptivní segmentací.

Subjekty 1, 2, 3		
Hledané shluky	Počet segmentů [-]	Proc. zastoupení [%]
Fyz. akt.	82722	90,3
Epilep.	5765	6,2
Pomal oč.	539	0,6
EMG	1972	2,2
Impulzní	509	0,6
Rov. čáry	61	0,1
celkem	91568	

Tabulka 17: Specificita (SP), senzitivita (SZ) a prediktivní hodnota pozitivního testu (PPV) pro algoritmy k-means, DENCLUE neupravený na EEG záznam a DENCLUE upravený na EEG záznam. Vše je získáno pro shluky fyziologické aktivity mozku, epileptických grafoelementů, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu, impulzních artefaktů a rovných čar u subjektů 1, 2 a 3.

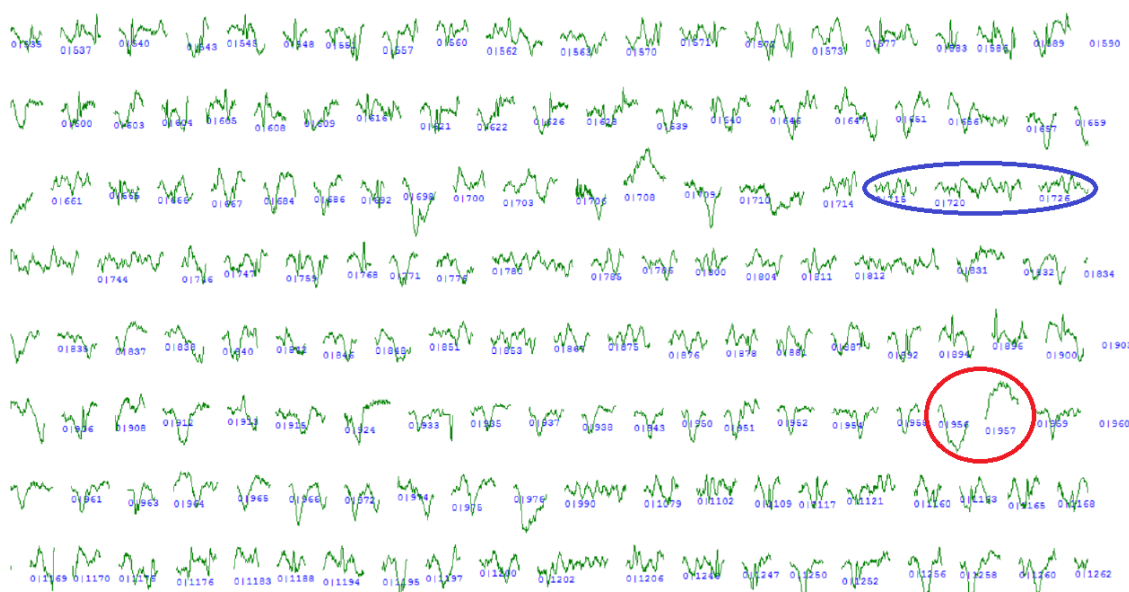
S. 1, 2, 3	k-means			neupr. DENCLUE			upr. DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	1,00	0,32	1,00	0,99	0,62	1,00	0,64	0,98	0,98
Epilep.	0,97	0,96	0,77	0,95	0,98	0,33	1,00	0,18	0,96
Pomal. oč.	0,87	0,74	0,01	alg. nenalezl shluk			1,00	0,00	0,21
EMG	1,00	0,96	1,00	1,00	0,00	0,10	0,97	0,82	0,35
Impulzní	1,00	0,48	1,00	alg. nenalezl shluk			1,00	0,45	1,00
Rov. čáry	0,97	0,18	0,01	1,00	0,56	1,00	1,00	0,98	0,84

V tabulce 16 lze vidět, že 90,3% segmentů z EEG záznamů subjektů 1 až 3 byly segmenty fyziologické aktivity mozku. Segmentů shluku epileptických grafoelementů bylo ve stejných EEG záznamech 5765. Segmentů shluku EMG artefaktů a zašuměného signálu bylo 1972. Segmentů rovných čar bylo u všech tří subjektů pouze 61.

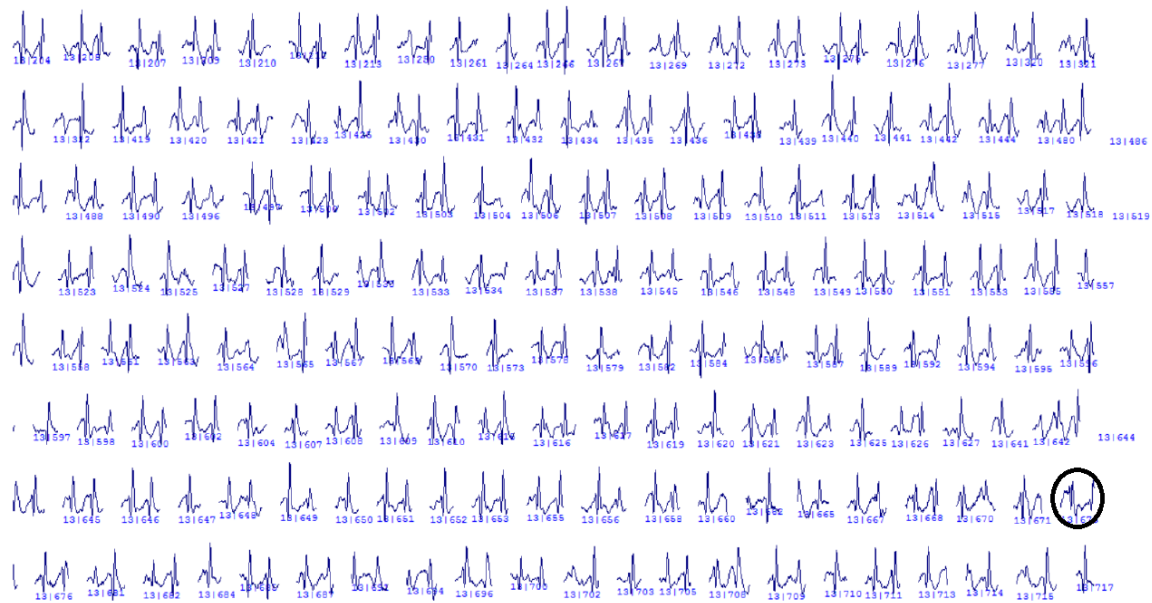
Z tabulky 17 lze vidět, že algoritmus DENCLUE neupravený na EEG záznam nenalezl ani u jednoho subjektu shluky pomalých očních artefaktů a impulzních artefaktů. Tento algoritmus měl u shluků rovných čar a fyziologické aktivity mozku PPV 1,00. Algoritmus k-means měl senzitivitu u shluku fyziologické aktivity mozku 0,32 a u shluku epileptických grafoelementů 0,96. Algoritmus DENCLUE upravený na EEG záznam měl

senzitivitu u shluku epileptických grafoelementů 0,18 a u shluku fyziologické aktivity mozku 0,98. U shluku fyziologické aktivity mozku měl algoritmus DENCLUE upravený na EEG záznam specificitu 0,64. Stejný algoritmus měl PPV u shluku EMG artefaktů a zašuměného signálu 0,35. Algoritmus k-means měl u shluku EMG artefaktů a zašuměného signálu PPV 1,00. Stejný algoritmus měl pro shluk pomalých očních artefaktů senzitivitu 0,74 a PPV 0,01.

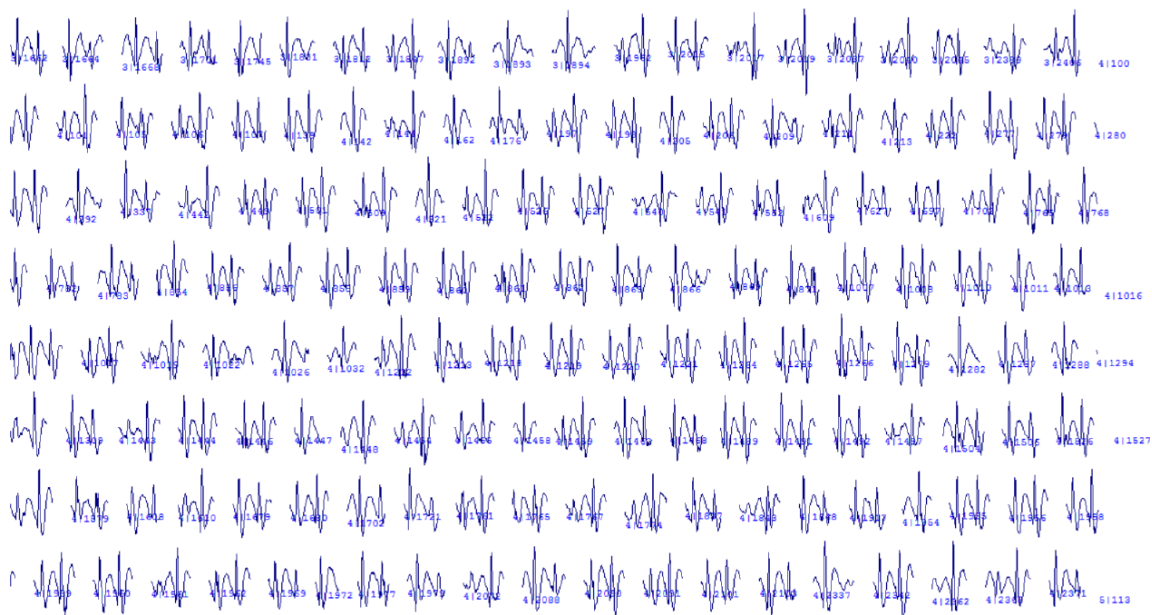
Na obrázcích 55 až 58 lze vidět příklady vzniklých shluků u subjektů 1 až 3 pro algoritmy k-means a DENCLUE upravený na EEG záznam. Obrázek 55 ukazuje příklad shluku fyziologické aktivity mozku vytvořeného algoritmem k-means. Na obrázku 56 lze vidět příklad shluku epileptických grafoelementů vytvořeného algoritmem k-means. Obrázek 57 ukazuje příklad shluku epileptických grafoelementů vytvořeného algoritmem DENCLUE upraveným na EEG záznam. Na obrázku 58 lze vidět příklad shluku fyziologické aktivity mozku vytvořeného algoritmem DENCLUE upraveným na EEG záznam.



Obrázek 55: Příklad shluku pomalých očních artefaktů vytvořeného algoritmem k-means. Příklad je vzat z EEG záznamů subjektů 1 až 3. Červeně je zde zakroužkován příklad očních artefaktů a modře je zakroužkován příklad fyziologické aktivity mozku, která do shluku nepatří. Zobrazeno v programu WF [5].



Obrázek 56: Příklad shluku epileptických grafoelementů vytvořených algoritmem k-means. Příklad je vzat z EEG záznamů subjektů 1 až 3. Černě je zde zakroužkovaný příklad špatné segmentace, která do shluku nepatří. Zobrazeno v programu WF [5].



Obrázek 57: Příklad shluku epileptických grafoelementů vytvořených algoritmem k-means. Všechny segmenty patří do shluku. Příklad je vzat z EEG záznamů subjektů 1 až 3. Zobrazeno v programu WF [5].



Obrázek 58: Příklad shluku fyziologické aktivity mozku vytvořeného algoritmem DENCLUE upraveným na EEG záznam. Příklad je vzat z EEG záznamů subjektů 1 až 3. Modře je zde zakroužkován příklad fyziologické aktivity mozku a červeně je zakroužkován příklad očních artefaktů, které do shluku nepatří. Zobrazeno v programu WF [5].

3.2.3 Kvalitativní analýza pro více EEG záznamů

Při kvalitativní analýze více EEG záznamů jsem hodnotil 12 EEG záznamů (od subjektů 4 až 15). U všech EEG záznamů jsem náhodně vybral 50 segmentů. Náhodný výběr jsem vytvářel za pomoci programového prostředí MATLAB R2015a [26]. Správnost zařazení vybraných segmentů do shluků jsem statisticky zpracoval. Pro statistické zpracování jsem použil ROC analýzu, konkrétně výpočet parametrů senzitivita, specifita a PPV. Statistické zpracování jsem použil pro porovnání algoritmů k-means a DENCLUE upraveného na EEG záznam. Tato část kvalitativní analýzy porovnála účinnost automatické klasifikace obou algoritmů pro více EEG záznamů.

V tabulce 18 lze vidět příklad rozložení 50-ti náhodně vybraných segmentů do hledaných shluků a jejich procentuální zastoupení ve výběru. Hodnoty představují správné zařazení segmentů do těchto shluků. Příklad v tabulce je ukázán pro subjekt 4. V tabulce 33 lze vidět specifitu, senzitivitu a PPV pro příklad stejného subjektu. Tyto parametry jsou ukázány pro správné shluky všech vybraných segmentů. Parametry jsem

počítal pro algoritmus k-means a algoritmus DENCLUE upravený na EEG záznam. Tabulky pro senzitivitu, specificitu a PPV u subjektů 5 až 15 pro algoritmus k-means a DENCLUE upravený na EEG záznam jsou ukázány v příloze.

Tabulka 18: Příklad počtu segmentů u každého hledaného shluku z 50 náhodných segmentů EEG záznamu subjektu 4. Procentuální zastoupení segmentů daných shluků v tomto výběru.

Subjekt 4		
Hledané shluky	Počet segmentů [-]	Proc. zastoupení [%]
Fyz. akt.	42	84,0
Epilep.	0	0,0
Pomal oč.	2	4,0
EMG	6	12,0
Impulzní	0	0,0
Rov. čáry	0	0,0
celkem	50	

Tabulka 19: Příklad hodnot specificity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 4. Hodnoty jsou získány pro shluky fyziologické aktivity mozku, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu.

Subjekt 4	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV [-]
Fyz. akt.	1,00	0,33	1,00	1,00	0,90	0,81
Pomal. oč.	1,00	1,00	0,33	1,00	1,00	0,04
EMG	0,98	0,33	1,00	0,98	0,33	0,02

Z tabulky 18 lze vidět, že náhodným výběrem nebyly u subjektu 4 nalezeny shluky epileptických grafoelementů, impulzních artefaktů a rovných čar. V tabulce 33 lze pozorovat, že shluk fyziologické aktivity měl senzitivitu 0,33 pro algoritmus k-means a 0,90 pro algoritmus DENCLUE upravený na EEG záznam. Tento algoritmus měl zároveň u shluku pomalých očních artefaktů PPV 0,04. Algoritmus k-means měl u stejného shluku PPV 0,33.

V tabulce 20 lze vidět součet rozložení náhodně vybraných segmentů do hledaných shluků a procentuální zastoupení těchto shluků ve výběru. Součet je proveden pro subjekty 4 až 15. Hodnoty představují správné zařazení segmentů do hledaných shluků. V ta-

bulce 21 lze vidět specificitu, senzitivitu a PPV pro stejné subjekty. Tyto hodnoty jsou ukázány pro správné shluky všech vybraných segmentů. Všechny hodnoty jsem počítal jako medián z hodnot subjektů 4 až 15. Hodnoty jsem počítal u algoritmu k-means a algoritmu DENCLUE upraveného na EEG záznam.

Tabulka 20: Počet segmentů u každého hledaného shluku 12 subjektů u nichž bylo náhodně vybráno 50 segmentů z EEG záznamu. Procentuální zastoupení segmentů daných shluků v tomto výběru.

12 subjektů		
Hledané shluky	Počet segmentů [-]	Proc. zastoupení [%]
Fyz. akt.	536	89,3
Epilep.	3	0,5
Pomal oč.	41	6,8
EMG	20	0,3
Impulzní	0	0,0
Rov. čáry	0	0,0
celkem	600	

Tabulka 21: Hodnoty specificity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Hodnoty jsou získány pro shluky fyziologické aktivity mozku, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu jako medián hodnot z 12 subjektů. U každého subjektu bylo náhodně vybráno 50 segmentů

12 subjektů	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV [-]
Fyz. akt.	1,00	0,32	1,00	1,00	0,90	0,81
Epilep.	0,50	1,00	0,50	0,75	1,00	0,08
Pomal. oč.	1,00	0,67	0,50	1,00	0,20	0,04
EMG	1,00	0,33	0,38	1,00	1,00	0,02

Z tabulky 20 je patrné, že náhodným výběrem jsem ani v jednom EEG záznamu nenalezl shluky impulzních artefaktů a rovných čar. U shluku epileptických grafoelementů jsem náhodným výběrem našel pouze 3 segmenty. Procentuální zastoupení segmentů shluku fyziologické aktivity v náhodném výběru bylo 89,3%.

V tabulce 21 si lze všimnout, že algoritmus k-means měl pro shluk fyziologické aktivity mozku senzitivitu 0,32 a PPV 1,00. Algoritmus DENCLUE upravený na EEG záznam měl

pro stejný shluk senzitivitu 0,90 a PPV 0,81. Algoritmus DENCLUE upravený na EEG záznam měl u shluku pomalých očních artefaktů PPV 0,04 a u shluku EMG artefaktů a zašuměného signálu PPV 0,02. Algoritmus k-means měl u shluku pomalých očních artefaktů PPV 0,50 a u shluku EMG artefaktů a zašuměného signálu PPV 0,38.

4 Diskuze

4.1 Simulovaná data

Pro simulovaná data jsem prováděl kvantitativní a kvalitativní analýzu. Kvantitativní analýzou jsem zjišťoval časovou náročnost algoritmu k-means a algoritmu DENCLUE neupraveného na EEG záznam. Časová náročnost udávala čas v sekundách, za který jednotlivé algoritmy dokončí svoji činnost. Časovou náročnost jsem zjišťoval pro různý počet segmentů simulovaných dat. Počet segmentů simulovaných dat se pohyboval od 164 segmentů do 3412 segmentů.

Udávaná hodnota časové náročnosti byla mediánem z 5-ti naměřených hodnot časové náročnosti. Z grafu na obrázku 42 lze vidět, že časová náročnost obou algoritmů stoupala s počtem segmentů dat. Pro nízký počet segmentů měl nižší časovou náročnost algoritmus DENCLUE. Ta ovšem stoupala výrazněji než časová náročnost algoritmu k-means. Pro větší počet segmentů měl tedy nižší časovou náročnost algoritmus k-means. Tato změna nastávala již okolo hodnoty 2200 segmentů. To je hodnota nižší než je počet segmentů v EEG záznamech.

Během kvalitativní analýzy simulovaných dat jsem využíval 4 typy simulovaných dat. Pro každý typ jsem určoval senzitivitu, specificitu a PPV. Tyto hodnoty jsem určoval i pro jednotlivé shluky simulovaných dat. Na všech simulovaných datech jsem porovnával algoritmus k-means a algoritmus DENCLUE neupravený na EEG záznam.

Prvním typem simulovaných dat byla data s výrazně oddělenými shluky. Pomocí nich jsem testoval schopnost algoritmů oddělit shluky, jejichž hranice jsou separovatelné. Z výsledků v tabulce 5 je patrné, že oba algoritmy oddělovaly tyto shluky s minimální chybou. Jediná hodnota, která nebyla pro celková data rovna jedné, byla senzitivita. Ta byla pro oba algoritmy rovna 0,99. To bylo způsobeno nezařazením jednoho až dvou objektů do správného shluku. Oba algoritmy nezařadily správně objekty, které byly od center shluků výrazněji vzdálené, jak lze vidět na obrázku 43.

Druhý typ simulovaných dat představovalo šest shluků vytvořených blízko u sebe. Těmito simulovanými daty jsem testoval schopnost algoritmů oddělit shluky, jenž se částečně prolínají. Z tabulky 6 lze vypožorovat, že lépe odděloval shluky algoritmus DEN-

CLUE, jenž měl všechny sledované hodnoty pro celá data vyšší než algoritmus k-means. Výraznější rozdíl u obou algoritmů byl především v parametru PPV, který byl pro algoritmus DENCLUE 0,98 a pro algoritmus k-means 0,79. Z toho plyne, že pokud algoritmus DENCLUE vybral určitý objekt jako součást shluku, tak daný objekt do tohoto shluku většinou patřil. Oba algoritmy měly hodnotu senzitivity okolo 0,70 pro celá simulovaná data. Z toho vyplývá, že oba algoritmy měly problém se zařazením všech objektů do jejich shluků. Algoritmus k-means vůbec nenalezl jeden shluk, namísto toho rozdělil jiný shluk na dva shluky. Toto bylo pravděpodobně způsobeno nevhodnými iniciačními centry, na jejichž pozici je algoritmus k-means citlivý (viz kapitola 1.2.2). Z obrázku 44 je patrné, že chyby algoritmu DENCLUE byly převážně způsobeny zařazením více objektů do šumového shluku. To by se dalo kompenzovat snížením koeficientu šumu ξ , což by ovšem vedlo k horšímu oddělování jednotlivých částečně prolnutých shluků.

Třetí typ simulovaných dat byl vytvořen ze tří shluků, které měly rozdílnou hustotu a byly od sebe různě vzdálené. Z tabulky 7 lze pozorovat, že algoritmus k-means zařadil veškeré objekty těchto dat bezchybně. Oproti tomu měl algoritmus DENCLUE především menší senzitivitu. Senzitivita algoritmu DENCLUE pro celá data byla 0,90. Z toho plyne, že algoritmus DENCLUE nezařadil do jednotlivých shluků všechny objekty, které tam patřily. Pokud se podíváme na obrázek 45 zjistíme, že chyba byla opět způsobena objekty klasifikovanými jako šum. Senzitivita byla pro algoritmus DENCLUE nejnižší u shluku číslo 3 a měla hodnotu 0,82. Tento shluk měl nejnižší hustotu a proto zde byl vliv koeficientu šumu nejvýraznější. Snížením koeficientu šumu bychom ovšem hůře rozeznali 2 shluky s malou vzájemnou vzdáleností.

Čtvrtý typ simulovaných dat obsahoval dva prolnuté shluky s různou hustotou. Tímto typem dat jsem testoval rozdíl mezi algoritmem na bázi vzdálenosti (k-means) a algoritmem na bázi hustoty (DENCLUE). Algoritmus na bázi hustoty by měl prolnuté shluky detekovat, algoritmus na bázi vzdálenosti nikoli. Z výsledků v tabulce 8 je patrné, že algoritmus DENCLUE opravdu oba shluky detekoval správně. Všechny tři sledované hodnoty byly pro celá data rovny jedné. U 2. shluku měla senzitivita hodnotu 0,98. Z obrázku 46 je patrné, že to způsobilo vytvoření menšího třetího shluku, který v datech obsažen ne-

byl. Algoritmus k-means oproti tomu rozdělil objekty zcela chybně. Hodnoty senzitivity, specificity a PPV se u něj pohybovaly mezi 0,58 a 0,59.

4.2 Reálné EEG záznamy

V dalším kroku jsem porovnával testované algoritmy na reálných EEG záznamech. Porovnání jsem prováděl na EEG záznamech naměřených v Nemocnici Na Bulovce pro 15 subjektů. K porovnání jsem využíval kvantitativní i kvalitativní analýzu testovaných algoritmů.

Během kvantitativní analýzy jsem porovnával časovou náročnost algoritmů k-means, DENCLUE neupraveného na EEG záznam a DENCLUE upraveného na EEG záznam. Časovou náročnost jsem počítal v závislosti na počtu segmentů EEG záznamu. Pro každý počet segmentů jsem získával výslednou časovou náročnost jako medián z 5 časových náročností. Hodnoty časové náročnosti jsou ovlivněny řadou vedlejších vlivů (např. výkon počítače se spuštěným algoritmem, spuštěné podprogramy na daném počítači atd.). Z tohoto důvodu není podstatná konkrétní hodnota časové náročnosti, ale její trend. Z tabulky 9 lze vyčíst, že algoritmus DENCLUE neupravený na EEG záznam měl nejnižší časovou náročnost 173,9 s a to pro 8957 segmentů. Oproti tomu měl algoritmus DENCLUE upravený na EEG záznam nejvyšší časovou náročnost 36,1 s. Algoritmus k-means měl nejvyšší časovou náročnost 25,8 s. V obou případech se jednalo o časovou náročnost při 41963 segmentech. Z toho vyplývá, že algoritmus DENCLUE neupravený na EEG záznam měl vyšší časovou náročnost než zbylé dva algoritmy. V obrázcích 47 a 48 lze vidět, že časová náročnost všech tří algoritmů stoupala s počtem segmentů EEG záznamu. Z obrázku 48 je zřejmé, že algoritmus k-means měl nižší časovou náročnost než algoritmus DENCLUE upravený na EEG záznam. Rozdíl mezi oběma algoritmy se zvětšoval s rostoucím počtem segmentů EEG záznamu. Výpočetní náročnost algoritmu DENCLUE upraveného na EEG záznam byla N^2 .

Během kvalitativní analýzy reálných EEG záznamů jsem využíval ROC analýzu. Konkrétně jsem využíval statistické parametry senzitivitu, specificitu a PPV. Tyto parametry jsem počítal pro každý hledaný shluk jednotlivých EEG záznamů. Hodnotu počítaných parametrů napříč EEG záznamy jsem získával jako medián. Pro obrovské množství seg-

mentů všech 15 EEG záznamů (307 438 segmentů) nemohli experti analyzovat celé EEG záznamy všech subjektů. Z toho důvodu jsem rozdělil kvalitativní analýzu na dvě části. V první části jsem analyzoval celé EEG záznamy pro subjekty 1 až 3. V druhé části jsem analyzoval náhodně vybrané segmenty u subjektů 4 až 15.

V první části kvalitativní analýzy EEG záznamů jsem porovnával algoritmy k-means, DENCLUE neupravený na EEG záznam a DENCLUE upravený na EEG záznam. Výsledky této části kvalitativní analýzy ukázaly účinnost testovaných algoritmů při rozřazování segmentů celého EEG záznamu do shluků. Algoritmy byly porovnávány s přihlédnutím na požadavky lékaře využívajícího automatickou klasifikaci EEG v klinické praxi.

Algoritmus DENCLUE neupravený na EEG záznam rozděloval EEG na dva velké shluky a několik drobných shluků. Jedním z velkých shluků byl shluk fyziologické aktivity mozku a druhým z velkým shluků byl šumový shluk. Ten obsahoval velké množství různých grafoelementů a artefaktů. Jedním z malých shluků byl shluk rovných čar. Jak lze vidět v tabulkách 11, 13 a 15 algoritmus DENCLUE neupravený na EEG záznam alespoň u jednoho subjektu nerozeznal shluky epileptických grafoelementů, EMG artefaktů a zašuměného signálu, pomalých očních artefaktů a impulzních artefaktů. Algoritmus měl u shluku fyziologické aktivity mozku PPV 1,00, což ukazuje na to, že algoritmus vybral do tohoto shluku segmenty do něj správně patřící. Z tabulky 17 lze pozorovat, že měl tento algoritmus PPV 1,00 i pro shluk rovných čar. Ačkoli algoritmus DENCLUE neupravený na EEG záznam dokázal rozpoznat shluk fyziologické aktivity mozku a shluk rovných čar, ostatní shluky nedokázal v každém EEG záznamu nalézt. Algoritmus DENCLUE neupravený na EEG záznam se tudíž nehodí na automatickou klasifikaci EEG záznamu, protože pro lékaře je mnohem důležitější nalézt shluk epileptických grafoelementů než shluk rovných čar.

Shluk rovných čar algoritmus k-means buď nenalezl, nebo našel, ovšem s velmi nízkým PPV a nízkou senzitivitou. Toto lze vidět v tabulkách 11 a 13. Ve stejných tabulkách lze vidět, že algoritmus DENCLUE upravený na EEG záznam měl pro shluk rovných čar nejnížší senzitivitu 0,96 a nejnížší PPV 0,68. Z těchto výsledků vyplývá, že algoritmus DENCLUE upravený na EEG záznam rozpoznal téměř všechny segmenty shluku rovných čar. Oproti tomu algoritmus k-means nebyl schopný shluk rovných čar rozeznat.

Pokud se shluk impulzních artefaktů vyskytoval v EEG záznamu zřídka, nenalezl ho ani algoritmus k-means ani algoritmus DENCLUE upravený na EEG záznam. Tyto výsledky lze pozorovat v tabulce 11. Z tabulky 13 je patrné, že pro častější výskyt shluku impulzních artefaktů měl algoritmus k-means senzitivitu 0,96, specificitu 1,00 a PPV 1,00. Algoritmus DENCLUE upravený na EEG záznam měl pro stejný případ senzitivitu 0,90. Oba algoritmy tudíž dobře rozeznávaly shluk impulzních artefaktů v případě, že byl obsažen v EEG záznamu ve velkém množství. V opačném případě ho často přiřazovaly do shluku EMG artefaktů a zašuměného signálu.

Pro shluk EMG artefaktů a zašuměného signálu měl algoritmus DENCLUE upravený na EEG záznam celkovou senzitivitu 0,82, jak lze vidět v tabulce 17. Pokud se v EEG záznamu vyskytoval nízký počet segmentů tohoto shluku, měl algoritmus DENCLUE upravený na EEG záznam nízkou PPV (viz tabulky 13 a 15). V opačném případě měl tentýž algoritmus PPV 0,84, jak lze vidět v tabulce 11. Z toho vyplývá, že algoritmus DENCLUE upravený na EEG záznam sice našel velké množství segmentů tohoto shluku, ovšem pokud se jich v celém záznamu nevyskytoval dostatek, zařazoval do tohoto shluku i drobné epileptické grafoelementy, fyziologickou aktivitu mozku a segmenty pulzního artefaktu. To lze pozorovat na obrázcích 50 a 54. Algoritmus k-means měl pro shluk EMG artefaktů a zašuměného signálu celkovou senzitivitu 0,96 a celkový PPV 1,00. Z toho plyne, že algoritmus k-means segmenty tohoto shluku většinou našel a správně oddělil od ostatních segmentů. To lze vidět na obrázku 53.

Shluk pomalých očních artefaktů algoritmus DENCLUE upravený na EEG záznam buď nenalezl, nebo našel s nízkou senzitivitou a specificitou. Z toho vyplývá, že tento algoritmus nerozeznával shluk pomalých očních artefaktů, místo toho přiřazoval segmenty tohoto shluku do shluku fyziologické aktivity mozku. To lze vidět na obrázku 58. Algoritmus k-means našel shluk pomalých očních artefaktů ve všech případech s celkovou senzitivitou 0,74 a specificitou 0,87. U všech subjektů pro něj byla nízká PPV, jak lze vidět v tabulkách 11, 13 a 15. Z výsledků vyplývá, že algoritmus k-means byl sice schopný rozeznat shluk pomalých očních artefaktů, ovšem nenalezl všechny artefakty a ve shluku se vyskytovalo velké množství segmentů patřících do shluku fyziologické aktivity mozku. To lze vidět na obrázku 55.

U shluku epileptických grafoelementů se senzitivita pro algoritmus k-means pohybovala mezi 0,71 a 0,91. Pro algoritmus DENCLUE upravený na EEG záznam se senzitivita pohybovala mezi 0,11 a 0,35. Celkové PPV měl tento algoritmus 0,96. Algoritmus k-means měl pro EEG záznamy s vyšším výskytem epileptických grafoelementů vysoké PPV (viz tabulky 13 a 15). Pro EEG záznam s malým množstvím epileptických grafoelementů měl algoritmus k-means PPV 0,05. Tyto výsledky dokazují, že algoritmus k-means dokázal vždy nalézt velké množství segmentů patřících do shluku epileptických grafoelementů. Počet nalezených segmentů bylo snižováno vytvořením dalšího shluku epileptických grafoelementů s velkou amplitudou. Při častém výskytu segmentů epileptických grafoelementů v záznamu zařazoval algoritmus k-means do shluku epileptických grafoelementů převážně správné segmenty, jak lze vidět na obrázku 56. V opačném případě se ve shluku vyskytovalo i velké množství segmentů fyziologické aktivity mozku (viz obrázek 49). Algoritmus DENCLUE upravený na EEG záznam dokázal najít pouze omezené množství epileptických grafoelementů. Díky tomu vytvářel malé shluky, které ovšem obsahovaly převážně epileptické grafoelementy. To lze vidět na obrázku 57. Oba algoritmy měly nižší pravděpodobnost nalezení epileptických grafoelementů o nízké amplitudě. Ty lze vidět na obrázcích 49 a 54.

U shluku fyziologické aktivity mozku měl algoritmus k-means PPV u všech EEG záznamů 1,00. Celkovou specificitu měl algoritmus k-means také 1,00. Senzitivitu měl algoritmus k-means v rozmezí od 0,28 po 0,49. Senzitivita se zmenšovala s rostoucím procentuálním zastoupením segmentů fyziologické aktivity mozku v EEG záznamu. Z výsledků vyplývá, že algoritmus k-means měl ve shluku fyziologické aktivity zařazené převážně správné segmenty. Algoritmus ovšem nedokázal nelézt všechny segmenty fyziologické aktivity mozku. To bylo způsobeno vytvářením více shluků, ve kterých byly dominantní segmenty fyziologické aktivity mozku. Algoritmus DENCLUE upravený na EEG záznam měl nižší specificitu u EEG záznamů, které nebyly výrazně zašuměné (viz tabulky 13 a 15). Pro zašuměný EEG záznam měl algoritmus specificitu 0,96. Celková senzitivita a PPV byly pro tento algoritmus u shluku fyziologické aktivity mozku 0,98. Algoritmus DENCLUE upravený na EEG záznam tedy dobře rozlišil většinu fyziologické aktivity

mozku. Pokud ovšem nebyl EEG záznam zašuměný, vyskytovaly se segmenty fyziologické aktivity mozku ve větším množství i v ostatních shlucích.

V druhé části kvalitativního hodnocení EEG záznamů jsem náhodným výběrem vybral 600 segmentů z 12 EEG záznamů. Tuto část kvalitativní analýzy jsem prováděl pro algoritmy k-means a DENCLUE upravený na EEG záznam. Algoritmus DENCLUE neupravený na EEG záznam jsem nepoužíval pro jeho neúčinnost na automatickou klasifikaci EEG záznamu.

Z tabulky 20 je patrné, že jsem náhodným výběrem nenašel ani jeden segment od shluků rovných čar a pulzního artefaktu. U shluku epileptických grafoelementů jsem našel pouze 3 segmenty. Pro tento shluk jsem tedy nezískal dostatečný počet údajů pro relevantní statistické zpracování. Shluk EMG artefaktů a zašuměného signálu byl v náhodném výběru zastoupen pouze ve 3%. Jeho výsledky tudíž zcela neodpovídali výsledkům z kvalitativní analýzy celých EEG záznamů, kdy nejmenší hodnota PPV u algoritmu k-means byla 0,96. Oproti tomu náhodným výběrem jsem získal pro stejný shluk u algoritmu k-means hodnotu PPV 0,38. Stejně jako u celého EEG záznamu měl shluk pomalých očních artefaktů nižší PPV pro oba porovnávané algoritmy. Shluk fyziologické aktivity mozku byl v náhodném výběru obsažen v 89,3%. V celých EEG záznamech byl obsažen v 90,3%. Z toho plyne, že špatné nacházení většiny shluků v náhodném výběru bylo způsobeno velkým množstvím segmentů shluku fyziologické aktivity mozku v poměru k ostatním segmentům. Pro shluk fyziologické aktivity mozku odpovídaly hodnoty z náhodného výběru hodnotám získaným v celých EEG záznamech. Senzitivita byla výrazně nižší u algoritmu k-means než u algoritmu DENCLUE upraveného na EEG záznam. Naopak PPV bylo pro tento shluk vyšší u algoritmu k-means.

Pro lékaře v klinické praxi je důležité najít shluky epileptických grafoelementů či očních a EMG artefaktů. Shluk fyziologické aktivity mozku je v celých EEG záznamech obsažen v 90,3%. Není tudíž důležité najít tento shluk, ale najít všechny ostatní shluky ve velkém množství segmentů fyziologické aktivity mozku. Algoritmus DENCLUE vykazoval velkou senzitivitu u shluků fyziologické aktivity mozku, rovných čar či impulzních artefaktů. Pro ostatní shluky měl ovšem nižší senzitivitu než algoritmus k-means. Algoritmus k-means měl oproti tomu nižší senzitivitu na shluky fyziologické aktivity mozku a rovných čar.

Z těchto poznatků je patrné, že algoritmus DENCLUE upravený na EEG záznam sice dokázal klasifikovat některé shluky EEG záznamů lépe než algoritmus k-means, ovšem shluky mající pro lékaře vyšší hladinu významnosti klasifikoval lépe algoritmus k-means. V současné době se tedy do klinické praxe více hodí algoritmus k-means, ovšem vylepšení algoritmu DENCLUE upraveného na EEG záznam by mohlo být možné při extrahování příznaků zaměřených na lepší klasifikaci pro lékaře důležitějších shluků.

5 Závěr

V této práci jsem se zabýval automatickou klasifikací EEG záznamu za pomoci algoritmu DENCLUE. Algoritmus DENCLUE jsem nejprve naprogramoval v programovém prostředí MATLAB. Následně jsem pro jeho spuštění a načtení analyzovaných dat vytvořil uživatelsky přívětivé prostředí. Dále jsem algoritmus DENCLUE implementoval na automatickou klasifikaci EEG signálu a zjistil jeho výpočetní náročnost. Algoritmus DENCLUE jsem následně porovnával s algoritmem k-means používaným v klinické praxi. Porovnání jsem prováděl pro simulovaná data a reálné EEG záznamy. U obou případů jsem provedl kvantitativní a kvalitativní zhodnocení. Při kvantitativní analýze jsem zjišťoval časovou náročnost porovnávaných algoritmů. U kvalitativní analýzy jsem statisticky vyhodnotil účinnost algoritmů DENCLUE a k-means na automatickou klasifikaci EEG záznamů a simulovaných dat. Díky tomu jsem splnil veškeré body zadání mé diplomové práce.

Časová náročnost pro všechny testované algoritmy rostla s rostoucím počtem segmentů. Algoritmus DENCLUE neupravený na EEG záznam měl časovou náročnost přibližně o 2 řády horší než zbylé dva algoritmy. Algoritmus DENCLUE upravený na EEG záznam měl pro nižší počet segmentů (řádově stovky segmentů) podobnou časovou náročnost jako algoritmus k-means. Trend časové náročnosti byl u algoritmu k-means méně strmý než u algoritmu DENCLUE upraveného na EEG záznam.

Algoritmus DENCLUE neupravený na EEG záznam nedokázal rozpoznat všechny pro lékaře podstatné shluky a není proto vhodný pro automatickou klasifikaci EEG záznamu. Algoritmy k-means a DENCLUE upravený na EEG záznam dokázaly dobře rozpoznat část shluků a část shluků rozpoznat nedokázaly nebo při jejich klasifikaci chybovaly. Pro lékaře je přitom v klinické praxi podstatnější najít v EEG záznamu místa s výskytem epileptických grafoelementů, pomalých očních artefaktů či EMG artefaktů. V tomto ohledu vykazoval lepší výsledky algoritmus k-means. Algoritmus DENCLUE upravený na EEG záznam měl oproti němu horší především rozpoznávání shluku epileptických grafoelementů.

Tato práce tedy ukázala, že algoritmus DENCLUE není lepší než algoritmus k-means pro automatickou klasifikaci reálných EEG záznamů. Je ovšem lepší pro automatickou

klasifikaci simulovaných dat. Pro zlepšení algoritmu DENCLUE u automatické klasifikace EEG záznamů by mohla pomoci optimalizace použitých příznaků.

Reference

- [1] William F Ganong. *Přehled lékařské fyziologie*. Galén, Praha, 20. edition, c2005.
- [2] Zdeněk Ambler. *Základy neurologie*. Galén, Praha, 7. vyd. edition, c2011.
- [3] Vladimír Krajča, Svojmil Petránek, Ivana Patáková, and Alpo Värri. Automatic identification of significant graphoelements in multichannel eeg recordings by adaptive segmentation and fuzzy clustering. *International Journal of Bio-Medical Computing*, vol. 28(1-2):71–89, 1991.
- [4] Emilio Parisotto, Youness Aliyari Ghassabeh, Siamak Freydoonnejad, and Frank Rudzicz. Eeg dimensionality reduction in automatic identification of synonymy. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 847–851, Toronto ON Canada, 2015. IEEE.
- [5] V. Krajča and S. Petránek. ”wave-finder”: a new system for an automatic processing of long-term eeg recordings. *Quantitative EEG Analysis - Clinical Utility and New Methods*, pages 103–106, 1993.
- [6] Georgiy R. Minasyan, John B. Chatten, Martha J. Chatten, and Richard N. Harner. Patient-specific early seizure detection from scalp electroencephalogram. *Journal of Clinical Neurophysiology*, vol. 27(issue 3):163–178, 2010.
- [7] Larbi Boubchir, Somaya Al-Maadeed, and Ahmed Bouridane. Effectiveness of combined time-frequency image and signal-based features for improving the detection and classification of epileptic seizure activities in eeg signals. In *2014 International Conference on Control, Decision and Information Technologies (CoDIT)*, pages 673–678, Metz, 2014. IEEE.
- [8] Hao Qu and J. Gotman. A patient-specific algorithm for the detection of seizure onset in long-term eeg monitoring. *IEEE Transactions on Biomedical Engineering*, vol. 44(issue 2):115–122, 1997.
- [9] Václav Gerla. Automated analysis of long-term eeg signals, 2012.

- [10] Josef Faber. *Elektroencefalografie a psychofyziologie*. ISV, Praha, vyd. 1. edition, 2001.
- [11] Yu-Chen Song, M. J. O'Grady, G. M. P. O'Hare, and Wei Wang. A clustering algorithm incorporating density and direction. In *2008 International Conference on Computational Intelligence for Modelling Control & Automation*, pages 722–725, Vienna, 2008. IEEE.
- [12] Alexander Hinneburg and Daniel A. Keim. A general approach to clustering in large databases with noise. *Knowledge and Information Systems*, vol. 5(issue 4):387–415, 2003-11-1.
- [13] M Berthold, John Shawe-Taylor, and Nada Lavrač. *Advances in intelligent data analysis VII*. Springer, Berlin, c2007.
- [14] Jitka Mohylová and Vladimír Krajčea. *Zpracování biologických signálů*. Vysoká škola báňská - Technická univerzita, Ostrava, 2007 edition, 2007.
- [15] Saeid Sanei and Jonathon A Chambers. *EEG signal processing*. John Wiley & Sons, Hoboken, NJ, 2007 edition, 2007.
- [16] Vladimír Krajčea and Jitka Mohylová. *Číslíkové zpracování neurofyziologických signálů*. České vysoké učení technické, V Praze, 1. vyd. edition, 2011.
- [17] Sana Tmar-Ben Hamida, Beena Ahmed, and Thomas Penzel. A novel insomnia identification method based on hjorth parameters. In *2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pages 548–552, Abu Dhabi, 2015. IEEE.
- [18] Sueli A. Mingoti and Joab O. Lima. Comparing som neural network with fuzzy c-means, k-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research*, vol. 174(issue 3):1742–1759, 2006.
- [19] China; Huazhong University of Science and Technology sponsored by Huazhong Normal University. *Education Technology and Computer Science (ETCS), 2010 Second International Workshop on*. IEEE, Piscataway, 2. edition, 2010.

- [20] Umut Orhan, Mahmut Hekim, and Mahmut Ozer. Eeg signals classification using the k-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications*, vol. 38(issue 10):13475–13481, 2011.
- [21] Nikita Gurudath and H. Bryan Riley. Drowsy driving detection by eeg analysis using wavelet transform and k-means clustering. *Procedia Computer Science*, vol. 34(34):400–409, 2014.
- [22] Paschalis A. Bizopoulos, Dimitrios G. Tsalikakis, Alexandros T. Tzallas, Dimitrios D. Koutsouris, and Dimitrios I. Fotiadis. Eeg epileptic seizure detection using k-means clustering and marginal spectrum based on ensemble empirical mode decomposition. In *13th IEEE International Conference on BioInformatics and BioEngineering*, pages 1–4, Chania, 2013. IEEE.
- [23] Salih Güneş, Kemal Polat, and Şebnem Yosunkaya. Efficient sleep stage recognition system based on eeg signal using k-means clustering based feature weighting. *Expert Systems with Applications*, vol. 37(issue 12):7922–7928, 2010.
- [24] Guojun Shi, Bingkun Gao, and Li Zhang. The optimized k-means algorithms for improving randomly-initialed midpoints. In *Proceedings of 2013 2nd International Conference on Measurement, Information and Control*, pages 1212–1216, Harbin, 2013. IEEE.
- [25] Eyup Cinar and Ferat Sahin. New classification techniques for electroencephalogram (eeg) signals and a real-time eeg control of a robot. *Neural Computing and Applications*, vol. 22(issue 1):29–39, 2013.
- [26] MATLAB®. The MathWorks, Inc., 2014. Release: 2014b.
- [27] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, vol. 27(issue 8):861–874, 2006.
- [28] Christian O’Reilly and Tore Nielsen. Revisiting the roc curve for diagnostic applications with an unbalanced class distribution. In *2013 8th International Workshop on*

- Systems, Signal Processing and their Applications (WoSSPA)*, pages 413–420. IEEE, 2013.
- [29] Christian O’Reilly and Tore Nielsen. Revisiting the roc curve for diagnostic applications with an unbalanced class distribution. In *2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA)*, pages 413–420. IEEE, 2013.
- [30] A. T. Tzallas, C. D. Katsis, P. S. Karvelis, D. I. Fotiadis, S. Konitsiotis, and S. Giannopoulos. Classification of transient events in eeg recordings. *Proceedings of the 2nd International Conference on Advances in Biomedical Signal and Information Processing*, 2nd:7, 2004.

A Kvalitativní analýza pro více EEG záznamů

Tabulka 22: Příklad hodnot specifity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 5. Hodnoty jsou získány pro shluky fyziologické aktivity mozku, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu.

Subjekt 5	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	1,00	0,28	1,00	1,00	0,79	0,81
Pomal. oč.	1,00	1,00	1,00	1,00	1,00	0,04
EMG	0,98	1,00	0,50	0,98	1,00	0,02

Tabulka 23: Příklad hodnot specifity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 6. Hodnoty jsou získány pro shluky fyziologické aktivity mozku a pomalých očních artefaktů.

Subjekt 6	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	1,00	0,33	1,00	1,00	1,00	0,81
Pomal. oč.	0,98	0,50	0,50	0,98	0,50	0,02

Tabulka 24: Příklad hodnot specifity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 7. Hodnoty jsou získány pro shluky fyziologické aktivity mozku, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu.

Subjekt 7	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	1,00	0,23	0,12	1,00	0,64	0,81
Pomal. oč.	1,00	0,67	0,12	1,00	0,00	0,04
EMG	0,98	0,25	0,12	0,98	1,00	0,02

Tabulka 25: Příklad hodnot specifcity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 8. Hodnoty jsou získány pro shluky fyziologické aktivity mozku, epileptických grafoelementů, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu.

Subjekt 8	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	0,93	0,36	1,00	0,93	1,00	0,13
Epilep.	0,93	0,00	0,00	0,96	0,50	0,09
Pomal. oč.	0,96	0,80	0,67	0,93	0,20	0,13
EMG	0,93	1,00	0,25	0,93	1,00	0,13

Tabulka 26: Příklad hodnot specifcity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 9. Hodnoty jsou získány pro shluky fyziologické aktivity mozku, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu.

Subjekt 9	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	0,99	0,35	1,00	0,99	0,90	0,16
Pomal. oč.	0,99	1,00	0,50	0,99	1,00	0,16
EMG	0,99	1,00	1,00	0,99	1,00	0,16

Tabulka 27: Příklad hodnot specifcity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 10. Hodnoty jsou získány pro shluky fyziologické aktivity mozku a pomalých očních artefaktů.

Subjekt 10	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	1,00	0,31	1,00	1,00	0,84	0,81
Pomal. oč.	1,00	0,60	0,75	1,00	0,40	0,04

Tabulka 28: Příklad hodnot specificity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 11. Hodnoty jsou získány pro shluky fyziologické aktivity mozku, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu.

Subjekt 11	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	1,00	0,41	0,93	1,00	0,90	0,81
Pomal. oč.	1,00	0,33	1,00	1,00	0,00	0,04
EMG	0,98	0,33	0,17	0,98	0,33	0,02

Tabulka 29: Příklad hodnot specificity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 12. Hodnoty jsou získány pro shluk fyziologické aktivity mozku.

Subjekt 12	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	1,00	0,36	0,32	1,00	0,96	0,81

Tabulka 30: Příklad hodnot specificity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 13. Hodnoty jsou získány pro shluky fyziologické aktivity mozku a pomalých očních artefaktů.

Subjekt 13	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	1,00	0,27	1,00	1,00	0,85	0,81
Pomal. oč.	1,00	1,00	0,50	1,00	1,00	0,04

Tabulka 31: Příklad hodnot specificity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 14. Hodnoty jsou získány pro shluky fyziologické aktivity mozku, epileptických grafoelementů, pomalých očních artefaktů, EMG artefaktů a zašuměného signálu.

Subjekt 14	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	1,00	0,30	1,00	1,00	0,95	0,81
Epilep.	0,91	1,00	1,00	0,91	1,00	0,07
Pomal. oč.	0,91	0,60	0,43	0,91	0,40	0,07
EMG	0,98	0,00	0,00	0,98	0,00	0,02

Tabulka 32: Příklad hodnot specificity (SP), senzitivity (SZ) a prediktivní hodnoty pozitivního testu (PPV) pro algoritmy k-means a DENCLUE upraveného na EEG záznam. Příklad je pro náhodný výběr 50 segmentů z EEG záznamu jednoho subjektu, konkrétně subjektu 15. Hodnoty jsou získány pro shluky fyziologické aktivity mozku a pomalých očních artefaktů.

Subjekt 15	k-means			DENCLUE		
	SP [-]	SZ [-]	PPV [-]	SP [-]	SZ [-]	PPV[-]
Fyz. akt.	1,00	0,26	1,00	1,00	0,66	0,81
Pomal. oč.	1,00	0,67	1,00	1,00	0,00	0,04

B Seznam příloh na CD

Tabulka 33: Seznam příloh, které jsou k této práci dodány na CD.

Seznam příloh na CD
Naprogramované funkce v programu MATLAB R2015a, včetně GUI
Soubory formátu .nra pro každý používaný EEG záznam (subjekt)